

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO.
ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
CAMPUS IZTACALA.

**Identificación, descripción y análisis de regiones
conservadas en la estructura primaria de las proteínas
de la familia cpn60 (chaperoninas): Implicaciones
estructurales, funcionales y evolutivas.**

T E S I S P R O F E S I O N A L

QUE PARA OBTENER EL TITULO DE:

B I O L O G O

PRESENTA:

MIGUEL RANGEL AGUILERA

TLANEPANTLA EDO DE MEXICO.

1995



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

MI MAS PROFUNDO Y SINCERO AGRADECIMIENTO
AL M en C VICTOR.M .VALDEZ LOPES, YA QUE SIN SU
VALIOSA Y DESINTERESADA COLABORACION NO HUBIERA
SIDO POSIBLE LA REALIZACION DE ESTE TRABAJO.

GRACIAS NO SOLO POR BRINDARME SU AYUDA
PROFESIONAL, SINO TAMBIEN POR COMPARTIR
SUS AFICIONES Y AMISTAD.



AL BIOLOGO ANTONIO LAZCANO- ARAUJO. R.
MI AGRADECIMIENTO Y ADMIRACION A UN GRAN
PROFESIONISTA. QUIEN HA SIDO CATALOGADO COMO EL
MEJOR EN SU AREA EN IBEROAMERICA.

A TODOS LOS INTEGRANTES DEL LABORATORIO
DE MICROBIOLOGIA DE LA FACULTAD DE CIENCIAS
AL DR LUIS MEDRANO, A LAS BIOLOGAS ANA MARIA
VELAZCO, RAQUEL ORTEGA Y AL BIOLOGO ARTURO
BECERRA. GRACIAS POR SU INVALUABLE AYUDA.

A LOS SINODALES REVISORES DE ESTE TRABAJO.
M en C VICTOR RAMON MORENO.
BIOLOGO RAFAEL QUINTANAR
BIOLOGA IRMA. E. DUEÑAS
BIOLOGO ELIAS PIEDRA IBARRA

GRACIAS POR SU TIEMPO , SU DEDICACION Y
SUS ATINADOS CONSEJOS

A TODOS MIS COMPAÑEROS CONDISUPULOS
QUE ALGUNA VEZ COMENZAMOS ESTA AVENTURA.

A TODOS LOS QUE AUN CONSERVAN LA FE Y LA
ESPERANZA.

A LA ESCUELA NACIONAL DE ESTUDIOS PROFESIONALES
IZTACALA.

GRACIAS POR HABERME SOPORTADO TANTO TIEMPO.

A MI ESPOSA SUSANA GRACIAS POR SU CARIÑO, PACIENCIA
Y COMPRESION.

A MI HIJO CESAR AUGUSTO POR SER FUENTE DE
INSPIRACION.

A MI PEQUEÑO HIJO QUE AUN NO LLEGA, PERO QUE AL
LLEGAR ME DARA MAS FUERZAS PARA CONTINUAR.

A TODA MI FAMILIA . GRACIAS A CADA UNO DE
ELLOS.

A LA UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO.

GRACIAS POR LA OPORTUNIDAD INMERECDAMENTE RECIBIDA,
GRACIAS POR BRINDAR OPORTUNIDAD A TANTOS JOVENES QUE
COMO YO ALGUNA VEZ SOÑAMOS CON SER ALGUIEN EN TUS AULAS.
GRACIAS POR RESCATAR DE LA MARGINACION Y LA IGNORANCIA A
TUS DISIPULOS QUE A TI ACUDEN. GRACIAS POR FORMAR A TANTAS
GENERACIONES VALIOSAS DE MEXICANOS.Y POR QUE GRACIAS A TI
SEGUIMOS TENIENDO FE EN EL FUTURO. Y POR QUE A PESAR DE QUE
SE TE ATAQUE INJUSTIFICADAMENTE HAS REPRESENTADO ,
REPRESENTAS Y REPRESENTARAS SIEMPRE LA ESPERANZA DE MILES
DE JOVENES.

VIVA LA UNIVERSIDAD.

INDICE DE TEMAS.

1. Resumen	3
2. Introducción.....	5
2.1 Definición de proteínas chaperonas.....	5
2.2 Familia hsp70.....	7
2.3 Familia hsp90.....	7
2.4 Familia hsp60 (chaperoninas).....	8
Características generales	
Tipo de chaperoninas.....	10
a) Chaperoninas citoplasmáticas (Tric).....	10
b) Chaperoninas bacterianas (GroEL).....	11
c) Chaperonina de cloroplasto (RUBP).....	14
d) Chaperoninas mitocondriales	16
(Hsp60, Proteína p1).	
2.5 Sistemática molecular.....	19
2.6 Concepto de homología y	20
la alineación de secuencias.	
3. Antecedentes.....	23
4. Objetivos.....	25
5. Material y métodos.....	26
6. Resultados.....	36
7. Discusión de resultados.....	68
8. Conclusiones.....	78
9. Referencias.....	80
10. Apéndice.....	92

RESUMEN

Las chaperoninas son proteínas que forman complejos oligoméricos de 14 subunidades idénticas, de peso molecular de 60 Kda cada una. Presentan múltiples funciones, y la principal de ellas es su habilidad para reconocer y modular los estados de plegamiento dentro de las células. Se encuentran en todos los organismos donde han sido buscados, desde bacteria hasta mamíferos. En estos últimos se localizan en los organelos, lo que sugiere que estas proteínas son de origen endosimbótico. Aquí se presentan resultados que son congruentes con esta hipótesis.

Desde el punto de vista evolutivo, las características funcionales de las chaperonas moleculares en general y, en particular, la familia de cpn60 (incluyendo el involucramiento de esta última en la respuesta a choque térmico), las hacen de vital importancia para los procesos metabólicos tempranos, por lo que es probable que en ello se finque su alto grado de conservación.

El péptido ha sido observado al microscopio electrónico y a partir de ello se ha descrito su estructura cuaternaria. Aún así, pocos hechos son claros a nivel de estructura secundaria y terciaria; lo único concluyente son los estudios de dicroísmo circular, de donde se deduce que este péptido posee al menos 60% de alfas hélices en su estructura secundaria.

Los resultados obtenidos en este trabajo en la predicción de estructura secundaria usando varios métodos en el programa PcGene son cercanos a esa cifra (50.6%), y dado que estos caracteres se alternan con las β plegadas, la cpn60 se clasificaría como alfa/ β ta

En este trabajo se identificaron y describieron 5 regiones altamente conservadas en la estructura primaria de la cpn60. Como resultado, se proponen algunas implicaciones estructurales, funcionales y evolutivas para una de estas zonas. Como estrategia para dilucidar estas regiones, se utilizaron varios recursos, que incluyen desde alineamientos múltiples y pareados, gráficas manuales, hasta logaritmos por ordenador, incluyendo criterios fisicoquímicos (como el promedio de hidropatía), además de un esbozo de estructura secundaria promedio.

En base a lo anterior se comprobó la homología de estas proteínas y se obtuvo un árbol filogenético, donde se pueden ver en forma muy clara las relaciones filogenéticas de los diferentes grupos, ya que estos se alinean de manera natural.

A partir de características muy conservadas en este péptido, en el presente estudio se propone un posible sitio catalítico, que estaría en la cuarta región descrita. Sin embargo, la relevancia de las zonas restantes, no es aun clara, aunque la posibilidad que se identifique un papel igualmente importante no se puede descartar.

2 INTRODUCCION

CHAPERONAS MOLECULARES.

2.1 Definición.

Son proteínas que influyen en el plegamiento de otras proteínas y no son componentes de la estructura funcional final de estas últimas (Hemmingsen *et al.* 1988). Esta es una definición muy general, ya que ahora muchas proteínas son conocidas como moléculas chaperonas, como por ejemplo, nucleoplasmina, (Para la cual Laskey (1978) acuñó el termino), hsp70, partículas de reconocimiento, Sec B y otras (Ang *et al.*1991; Ellis y Van der vies, 1991). Las diferentes clases de chaperonas se distinguen en función de su peso molecular.

Las chaperonas, son llamadas también proteínas de choque térmico o de estrés. Las funciones de esas proteínas bajo condiciones celulares normales así como condiciones de estrés son actualmente foco de gran interés.

Su función *in vivo* no es como catalizadora de formación de estructura secundaria, sino de reorganización y estabilización de intermediarios no plegados o parcialmente plegados durante el proceso de plegamiento, ensamble y desensamble de los polipéptidos y, además previene la interacción intra e intercadenas que pueden provocar la formación de agregados.

Algunos miembros también interactúan con moléculas nativas para promover el reagrupamiento de las interacciones proteína-proteína en estructuras oligoméricas (Ellis y Hemmingsen, 1989; Gething y Sambrook, 1992).

La mayoría de las chaperonas identificadas hasta ahora corresponden a tres familias altamente conservadas de proteínas, cuyos miembros están ampliamente distribuidos desde procariontes hasta plantas y mamíferos. En células eucariontes los miembros de las diferentes familias han sido localizados en diferentes compartimientos y organelos celulares. Su principal denominador puede ser su acumulación en los casos donde aparezcan proteínas no plegadas o mal plegadas en la célula (Ananthan *et al.* 1986; Normington *et al.* 1989). Los nombres de estas familias son: hsp70, hsp90 y hsp60 (groEL/chaperoninas).

El interés inicial hacia estas familias de proteínas fue debido a su inducción específica durante la respuesta de los organismos al choque térmico (Lindquist y Craig, 1988). No obstante, la mayoría de los miembros de esas familias son expresados constitutiva y abundantemente, aun en ausencia de algún tipo de estrés. Además en diversos estudios genéticos se ha demostrado que muchas de esas proteínas son esenciales para la viabilidad celular bajo condiciones de crecimiento normal (Lindquist y Craig, 1988; Fayet *et al.* 1989).

2.2 La familia hsp70.

Los miembros de esta familia de proteínas han sido implicados en la estabilización de proteínas no plegadas antes de su ensamble en el citoplasma (Beckmann *et al.*1990), en la translocación dentro de los organelos incluyendo el retículo endoplásmico y mitocondria (Chirico *et al.*1988;Murakami *et al.*1988), en la estabilización de polipéptidos nuevos translocados antes de su plegamiento y ensamble (Pelham,1989;Kang *et al.*1990), en el rearreglo de proteínas oligoméricas (Rothman y Schmid,1986;DeLuca *et al.*1990), en la disolución de agregados proteicos (Pelham,1989), y en la degradación rápida de proteínas inviables en el citoplasma (Dice,1990;Chiang *et al.* 1989).

2.3 La familia hsp90.

Los miembros de este grupo quizás estabilicen una amplia variedad de proteínas inactivas o no plegadas (Lindquist y Craig,1988). constituyen el segundo grupo mayor de proteínas de estrés, sus miembros están ampliamente distribuidos. Las hsp90 en el citoplasma están asociadas con diversas proteínas celulares, incluyendo las retrovirales transformantes, los receptores hormonales (esteroides), las cinasas celulares, la actina y la tubulina (Lindquist y Craig,1988). Las características comunes de esas interacciones parece ser la estabilización de las proteínas en estado inactivo o no plegadas.

La función de las hsp90 no está aún clara, sin embargo es posible que tengan un papel similar a las proteínas hsp60 y hsp70, por ejemplo uniendo a los polipéptidos nacientes y acompañando a estas proteínas hasta su localización celular correcta (como en el caso de las hsp60), o manteniendo en ellas su forma correcta para el plegamiento.

2.4 Hsp60 (chaperoninas).

Las chaperoninas son una familia de proteínas altamente conservadas filogenéticamente (Ellis y Van der Vies, 1991; Hemmingsen *et al.* 1988; Zeilstra *et al.* 1991). Este péptido es codificado por el gen *groE* (Georgopoulos *et al.* 1973; Sternberg, 1973). Se ha demostrado por diversos experimentos que las proteínas de esta familia no solo son homólogas en secuencia, sino que funcionalmente también están altamente conservadas (Gatenby *et al.* 1985).

Su alto grado de identidad, así como su amplia distribución (se encuentran en células procariontes y eucariontes, en particular en los organelos de estos últimos) sugieren un origen muy antiguo, además de alguna relación con la aparición de mitocondrias y cloroplastos vía endosimbiosis.

En condiciones naturales son abundantes, y en diversas situaciones adversas (por ejemplo aumento de temperatura) se puede incrementar drásticamente su nivel de síntesis, por lo que se les ha incluido en las llamadas proteínas de choque térmico o de estrés, pero en forma errónea ya que ello solo alude a una de sus funciones.

Características generales

Son oligómeros de alto peso molecular, con 14 subunidades idénticas, cada una con una masa molecular aproximada de 60 Kda, arregladas en una estructura compleja que comprende 2 anillos de 7 subunidades cada uno, asociados uno a otro. Este tipo de molécula es conocida como GroEL (en bacterias), Hsp60 (en mitocondria de levadura), RUBP (en cloroplasto de planta) y proteína p1 (en mitocondria de mamíferos) o colectivamente como chaperoninas 60 (Cpn60) en razón del peso molecular de sus subunidades.

En contraste con las cpn60 de bacterias y de mitocondrias de levadura, las cuales contienen una subunidad simple (Hemmingsen *et al.* 1988; Reading *et al.* 1989), la proteína del cloroplasto contiene 2 polipéptidos de poca diferencia en tamaño (Martel *et al.* 1990; Musgrove *et al.* 1987) pero de alta divergencia en secuencia de aminoácidos. Sin embargo, existe controversia acerca de si la unidad funcional es homo o hetero oligomérica.

GroEL posee una actividad ATPasa débil, que requiere iones de magnesio y potasio para su máxima actividad (Chandrasekhar *et al.* 1986; Viitanen *et al.* 1990). Esta actividad es suprimida cuando GroES se une a GroEL, lo que requiere la presencia de otro ATP o ADP (Chandrasekhar *et al.* 1986; Viitanen *et al.* 1991).

Chaperoninas 10 (Co-chaperoninas cpn10).

Es una proteína pequeña, que contiene subunidades de cerca de 10 Kda, y al menos en bacteria parece estar formada de un anillo simple de 7 subunidades (Chandrasekhar *et al.* 1986). En bacterias es conocida como GroES, y las proteínas relacionadas en mitocondrias de mamífero son conocidas como cpn10 (Lubben *et al.* 1990).

Aunque se desconoce el papel biológico específico de cpn10, por evidencias experimentales se ha propuesto que normalmente funciona interactuando con la chaperonina cpn60, regulando su función (Chandrasekhar *et al.* 1986; Goloubinoff *et al.* 1989; Viitanen *et al.* 1991). Aunque con bajo nivel de resolución, las estructuras cuaternarias de las chaperoninas bacteriana y mitocondrial cpn60 han sido observadas al microscopio electrónico (Bochkareva *et al.* 1988; Lissin *et al.* 1990; McMullin *et al.* 1988). La estructura secundaria y terciaria de ambas chaperoninas (10 y 60 Kda) es aún desconocida.

Tipo de chaperoninas.

a). Chaperoninas citoplasmáticas.

Recientemente se ha demostrado con técnicas de microscopía electrónica y de inmunomarcaje que GroEL y las proteínas relacionadas están también presentes en el citoplasma de células de avena (Grimm *et al.* 1991). También se ha visto que algunas secuencias homólogas pueden ser identificadas entre las chaperoninas y el complejo polipeptídico citoplasmático 1 de ratón (tcp1) (Gupta, 1990). Sin embargo, su relación a nivel de secuencia con cpn60 es muy débil (17%), además de que presenta diferencias en su actividad de chaperona, y no presenta actividad de respuesta al choque térmico (Gupta, 1990).

Por lo anterior se le ha propuesto como una familia nueva, para la cual se ha sugerido el nombre de Tric (Hendrick y Hartl, 1993). A esta familia también pertenecería tf55 (factor termofílico de subunidades de 55 Kda), una proteína chaperona y con actividad de respuesta al choque térmico de *Sulfolobus shibatae*, además de otras proteínas relacionadas, todas ellas de arqueobacteria (Trent *et al.* 1991; Frydman *et al.* 1992; Gao *et al.* 1992; Lewis *et al.* 1992; Yaffe *et al.* 1992; Hendrick y Hartl, 1993).

Por otro lado, en el momento de elaborar este trabajo se tenían un número muy escaso de secuencias disponibles de este tipo de polipéptidos. Además de que los estudios sobre esta proteína están en pleno desarrollo, por lo que en el presente ensayo no se han incluido.

b) Las chaperoninas Bacterianas.

Las primeras chaperoninas estudiadas en detalle fueron las proteínas GroES y GroEL de *E. coli*. El gen *groE* fue originalmente identificado por mutaciones, que evitaban el crecimiento de algunos bacteriófagos (Ang *et al.* 1991; Ellis y Van der vies, 1991; Zeilstra *et al.* 1991). Después se descubrió que las proteínas *groE* influyen el ensamble de cabeza o estructura del tallo (dependiendo del fago). Los sitios de esa integración fueron identificados genéticamente, y estudios sucesivos revelaron que tanto los genes de GroES como los de GroEL son esenciales para el crecimiento bacteriano (Fayet *et al.* 1989), y que constituyen un operón cuya expresión es incrementada durante el choque térmico.

En condiciones de estrés moderado, el nivel de GroEL celular puede aumentar de un 2% a un 10% con respecto al nivel basal (Ellis y Van der vies, 1991). Evidencias genéticas sugieren que GroEL está también involucrada en la replicación de DNA, y en el ensamble y transporte de proteína de los fagos (Fayet *et al.* 1986; Goloubinoff *et al.* 1989; Jenkins *et al.* 1986; Van Dyk *et al.* 1989).

En otras bacterias también han sido identificadas proteínas relacionadas a GroES o GroEL, mismas que poseen un alto grado de similitud en sus secuencias de aminoácidos (Zeilstra *et al.* 1991). Secuencias homólogas también se han identificado entre *E. coli* y el cloroplasto *cpn60* (Hemmingsen *et al.* 1988). Esta última proteína está implicada en el ensamble de la Rubisco en cloroplasto (Gatenby y Ellis, 1990).

Se ha demostrado experimentalmente que es posible expresar y ensamblar Rubisco cianobacterial en *E. coli* (Gatenby *et al.* 1985), y que es factible la utilización de genética molecular bacteriana para probar si los productos del gen *groE* están involucrados en el ensamble *in vivo* de Rubisco.

Lo anterior sirvió de base para verificar el papel de las chaperoninas, y que fueron detectadas mediante incremento del ensamble de Rubisco, y en la utilización de cadenas de *groE* defectuosas para anticipar una reducción en el ensamble de Rubisco (Goloubinoff *et al.* 1989).

Se ha demostrado también experimentalmente el papel de GroEL en la regulación del gen *nif*, y en el ensamble de la Nitrogenasa (Govezensky *et al.* 1991).

En un intento por definir el papel de las chaperoninas en células bacterianas, se han utilizado técnicas de supresión genética, pues el plegamiento de polipéptidos puede estar influenciado significativamente por la temperatura, y se sospecha que en bacterias algunas mutaciones sensibles al calor pueden producir plegamientos alterados. En este tipo de incubación a temperaturas no permisivas, el calor propicia la desestabilización de intermediarios del plegamiento, dando por resultado la agregación y proteólisis, y en consecuencia un crecimiento defectuoso.

Se demostró así que el incremento de la expresión de las chaperoninas *groE*, en un rango de temperaturas que afectan al crecimiento de mutantes, puede corregir crecimientos defectuosos para el ensamble y/o plegamiento de polipéptidos mutantes.

También se ha encontrado que la sobreexpresión de los productos del operón *groE* dan como resultado la supresión de mutaciones sensibles al calor en genes que codifican diversas enzimas y proteínas estructurales (Van Dyk *et al.* 1989).

Una supresión similar también fue observada para los fenotipos mutantes de algunos alelos *dnaA* (Fayet *et al.* 1986; Jenkins, 1986), notándose que esta supresión requiere de la sobreexpresión de *groES* y *groEL*. Los niveles requeridos de las proteínas *groE* están alrededor del 20 al 30% del total de la proteína celular. Es posible que este fenómeno se deba al reacomodo de esas proteínas inestables mutantes. Es probable también que con niveles normales de chaperonina en la célula se pueda auxiliar a corregir errores en el plegamiento de proteínas mutadas, y de esta manera corregir algunos defectos genéticos que traen consigo plegamientos débiles.

Estas evidencias genéticas muestran que, en conclusión muchas proteínas en bacteria interactúan con chaperoninas, y lo cual ha sido comprobado con la demostración de su elevada afinidad de asociación a *groEL*.

c) Chaperoninas de cloroplastos.

Las chaperoninas cpn60 de cloroplastos fueron encontradas inicialmente estudiando la biosíntesis de Rubisco en cloroplastos aislados. Las cpn60 fueron vistas a través de geles de poliacrilamida, en donde se puede ver que la subunidad grande de Rubisco sintetizada en el cloroplasto se ensambla en la holoenzima, pero una parte importante de esa subunidad se asocia en forma estable con una proteína oligomérica de alto peso molecular (Barraclough y Ellis, 1980).

Este oligómero fue inicialmente llamado subunidad proteica grande de unión a Rubisco (conocida ahora como cpn60), y posee un tamaño de más de 800 Kda, con subunidades de 60 Kda (Barracough y Ellis, 1980; Hemmingsen y Ellis, 1986).

Estos estudios sugieren la posibilidad de que la subunidad grande de Rubisco naciente este específicamente asociada con cpn60 antes de su ensamble con la holoenzima, y que el complejo binario de la subunidad grande/cpn60 sea un intermediario esencial en el ensamble de Rubisco (Barracough y Ellis, 1980).

En experimentos independientes se ha demostrado que la subunidad grande de Rubisco sintetizada *in vivo* puede recuperarse intacta de cloroplastos en forma de dos complejos de sedimentación de 7s y 29s en gradiente de sacarosa (Roy *et al.* 1982). El complejo 29s contiene subunidad de Rubisco grande asociada con cpn60, y el complejo 7s puede comprender dímeros de Rubisco.

Cuando los cloroplastos fueron incubados a la luz, se observó que las nuevas subunidades grandes sintetizadas en el complejo 7s y en el 29s desaparecieron y son posteriormente encontradas en el ensamble de la holoenzima Rubisco, lo que demostró que la subunidad 29s interactúa con la cpn60 (Roy *et al.* 1982).

El ensamble postranslacional de Rubisco es acelerado en extractos de cloroplasto por la adición de ATP. Remanentes de complejo de 29s y oligómeros de cpn60 permanecen intactos (Blom *et al.* 1983).

Sin embargo en presencia de iones de Magnesio el ATP causa disociación de el complejo 29s (Bloom *et al.* 1983). Adicionalmente se vio que cpn60 purificado de cloroplasto fue utilizado exitosamente en el replegamiento de Rubisco *in vitro*, (Goloubinoff *et al.* 1989).

En cloroplasto, cpn60 está compuesto de dos tipos de subunidades de 61 Kda y 60 Kda, conocidos como subunidades alfa y beta (Musgrove *et al.* 1987) respectivamente. Las dos subunidades son altamente divergentes en su predicción de secuencia de aminoácidos (Martel *et al.* 1990). Ambas son codificadas en el núcleo por genes diferentes, y son importadas al cloroplasto a través de la síntesis de precursores de ribosomas citoplasmáticos (Hemmingsen y Ellis, 1986). La homología de la subunidad alfa de plantas con GroEL de *E. coli*, fue demostrada por Hemmingsen *et al.* (1988), alineando ambas secuencias.

Aunque está demostrado el papel de cpn60 del cloroplasto en el ensamble de Rubisco (Barraclough y Ellis, 1980; Bloom *et al.* 1983), esta tiene una función más general en la biogénesis del cloroplasto. En este sentido muchas proteínas importadas en el cloroplasto forman un complejo estable con el oligómero cpn60 (Gatenby *et al.* 1988; Lubben *et al.* 1989). Estos polipéptidos importados están asociados con cpn60, y pueden ser liberados de esta por la adición de magnesio/ATP.

En conclusión, al penetrar las proteínas al interior del cloroplasto en su estado no nativo, pueden requerir interacción con cpn60 para mediar su repliegamiento correctamente, como ha sido observado en mitocondria. Una similitud adicional es que el cloroplasto también contiene una co-chaperonina cpn10 homóloga, que aparentemente ayuda en la descarga eficiente del polipéptido asociado a cpn60 del plastido (Lubben *et al.* 1989).

d) Chaperoninas Mitocondriales.

Tanto la cpn60 y la cpn10, están presentes en mitocondria. La cpn60 fue caracterizada preliminarmente como una proteína que se acumulaba en la mitocondria de *Tetrahymena thermophila* durante el choque térmico. La proteína purificada de este organismo contiene subunidades de 58 Kda, y sedimenta un complejo de 20s a 25s en gradiente de sacarosa, en donde se obtiene una estructura relacionada a la proteína de GroEL de *E. coli*, y a proteínas cpn60 presentes en mitocondria de hongos, plantas y animales (McMullin *et al.* 1988).

Los análisis de genes clonados indican una identidad de aproximadamente 50% en la secuencia de aminoácidos entre las proteínas cpn60 de bacteria, mitocondria humana, de levadura y cloroplasto (Reading *et al.* 1989). Se sabe que la cpn60 mitocondrial está codificada en el núcleo, y que su síntesis se incrementa como respuesta al choque térmico (Reading *et al.* 1989).

En análisis de cepas silvestres y mutantes se demostró que la proteína cpn60 es requerida para el plegamiento correcto, y para el ensamble de proteínas importadas a mitocondria (Cheng *et al.* 1989).

Se pudo ver así, que cuando una cepa con mutaciones en el gen *groE*, es crecida a temperaturas no permisibles, el ensamble de un número importante de proteínas es defectuoso. Ello indica una disrupción en el estado oligomérico de *cpn60* a temperaturas no permisivas, lo que resulta en un fenotipo pleiotrópico no ensamblable (Ostermann *et al.* 1989).

En estudios posteriores se observó que, como sucede en cloroplasto, en ausencia de Mg/ATP una variedad de proteínas importadas en mitocondrias aisladas forman asociaciones estables con *cpn60* (Lubben *et al.* 1989), pero al agregar Mg/ATP la chaperonina libera a la proteína la cual adopta su conformación correcta.

Estos experimentos sugieren que la unión de las proteínas, se da al menos en parte en la superficie de la *cpn60*, o que esta las libera rápidamente (Martin *et al.* 1991).

Otros datos demuestran que las proteínas importadas de mitocondria pueden interaccionar inicialmente con chaperonas *hsp70*, y que al liberarse de *hsp70* el polipéptido interactúa con *cpn60*, completando así el plegamiento y ensamble (Manning-Krieg *et al.* 1991).

También existe una asociación específica entre *cpn60* y proteínas sintetizadas dentro de mitocondria. Por ejemplo, la subunidad alfa de F1/ATPasa sintetizada en mitocondria de maíz se une a la *cpn60* endógena y se libera por la adición de Mg/ATP, lo que demuestra que el plegamiento de proteínas codificadas tanto por el núcleo como por la mitocondria está mediado por chaperoninas (Prasad *et al.* 1990).

Como se ha descrito anteriormente, la tasa de polipéptidos plegados usualmente aumenta cuando el Mg/ATP, y la Co-chaperonina cpn10 están presentes juntas. Por lo tanto, no resulta sorprendente que una GroES homóloga este también presente en mitocondria aislada de mamíferos (Lubben *et al.* 1990).

La cpn10 y la GroEL mitocondrial forman un complejo estable en presencia de Mg/ATP, facilitando el aislamiento e identificación de la co-chaperonina (Lubben *et al.* 1990). Esta posibilidad también ha facilitado la identificación de co-chaperoninas en cloroplasto.

La cpn10 mitocondrial puede reemplazar a la GroES bacterial en el replegamiento dependiente de chaperoninas en la desnaturalización química de Rubisco, resultando la formación de un complejo con GroEL. Lo anterior demuestra como la maquinaria de plegamiento de procariontes y mitocondrias requieren de co-chaperoninas, para su máxima actividad biológica.

2.5 Sistemática Molecular

La sistemática molecular ha ido creciendo en importancia en los últimos años, ya que presenta ventajas sobre otro tipo de datos usados en la reconstrucción filogenética, como los morfológicos, los ontogenéticos, los etiológicos, etc.

Los datos moleculares tienen atributos únicos, como son la claridad y precisión con la que se definen los caracteres y los estados de estos caracteres.

En general, los caracteres son independientes y discretos, o sea, que, con diferentes tipos de datos moleculares es posible establecer homología con los nuevos taxa, de tal manera que los datos de diferentes estudios pueden no solo ser utilizados sino también combinados con otros, estos se vuelven aditivos. Obviamente se supone que estos caracteres están sujetos a el mismo proceso evolutivo. Además dado que ellos son homogéneos, el número de caracteres que se puede manejar es extremadamente grande. Una ventaja adicional es que estos se pueden comparar dentro de un amplio rango de taxa diferentes, y por lo tanto permiten hacer análisis que no se podrían hacer con otro tipo de información. Por ello los datos moleculares pueden proponer respuestas a todos los niveles de la jerarquía taxonómica.

2.6 El concepto de homología y la alineación de secuencias

La definición de homología de los caracteres es crítica en todo estudio filogenético. En general, se dice que los caracteres son homólogos cuando tienen un origen común. En sistemática molecular, el uso del termino "homología" ha adquirido algunos matices, lo que ha generado nuevos términos para enfatizar el sentido de homología en el contexto de la nomenclatura de secuencias de proteínas y de ácidos nucleicos.

Por ejemplo, **ortología** es la homología debida a la especiación; **Paralogía** es la homología debida a los eventos de duplicación de la secuencia antes de su especiación; **xenología** es la homología debida a los eventos de transferencia horizontal de material genético. En este sentido, es importante hacer notar que, si se tiene interés en estudiar las relaciones entre especies y no las relaciones entre los genes, se debe incluir solo secuencias ortólogas.

La alineación de las secuencias es un factor crucial para cumplir con el criterio de homología. En el proceso de alineación, las posiciones de los nucleótidos o aminoácidos son asignadas para cotejarlas solamente con sus equivalentes de otras secuencias.

En una alineación de dos o más secuencias, los "caracteres" corresponden a las posiciones de nucleótidos o aminoácidos en la secuencia. Los "estados" de estos caracteres son los diferentes nucleótidos o aminoácidos observados en esa posición.

En ocasiones no es tan sencillo obtener la alineación correcta de varias secuencias, especialmente en casos cuando se presenten deleciones e inserciones. Por ello, la alineación es quizás el paso más crítico y difícil de la sistemática molecular, y con frecuencia es el procedimiento menos comprendido en los análisis filogenéticos basados en datos de secuencias.

El procedimiento básico para alinear consiste en la asignación de valores positivos para los residuos alineables, a los residuos no alineables no se les asigna valor, y se asignan valores negativos para los "gaps" (espacios que resultan al realizar la alineación entre 2 o más secuencias). La alineación que maximiza el número obtenido es considerada como la mejor.

En una alineación de secuencias, los estados diferentes de un carácter corresponden a eventos de sustitución y los gaps corresponden a inserciones o deleciones. Los gaps son colocados cuando al insertarlos se maximiza el valor de similitud total.

Para la alineación de conjuntos de varias secuencias se han desarrollado diversos métodos, que se basan en diferentes suposiciones e intentan alinear a las secuencias bajo distinto criterio. Asimismo, existen también programas que simultáneamente realizan la alineación y la reconstrucción de filogenias.

Un ejemplo es el programa CLUSTAL (Higgins y Sharp, 1989) que funciona de esta manera. Si bien es cierto que existen diferentes programas de software que realizan la alineación, estos no siempre dan como resultado la alineación correcta, y frecuentemente es necesario realizar ajustes a ojo a las alineaciones realizadas por cómputo. La medida de "distancia" es un concepto muy utilizado en los estudios de sistemática molecular. La definición de distancia es muy ambigua. Usualmente la distancia es estimada como un sinónimo de disimilitud y opuesto a similitud. Los valores de disimilitud al igual que los de similitud son referidos en una escala que va de 0 a 1 o de 0 a 100%, y la disimilitud se define como la similitud menos 1.

En los análisis fenéticos los datos de secuencia son transformados a distancias por medio de un algoritmos de transformación. Existen diferentes tipos de modificaciones, cada uno basado en diferentes modelos. Un ejemplo es el método utilizado por Doolittle (1986), calculado a partir de los valores de similitud. Aun así, el problema de las distancias entre las secuencias es más complejo de lo que aparenta, puesto que el número de los estados de los caracteres puede ser mayor que el cambio observado, debido a la estructura de el código genético y/o por los eventos de sustitución múltiple.

En un análisis cladístico los datos de secuencia son generalmente tratados como un multiestado de caracteres, donde el número de los estados del carácter es usualmente 4 para DNA y 20 para aminoácidos. Sin embargo, se pueden incorporar otras suposiciones a estos análisis y reducir así la cantidad de estados de los caracteres.

Cuando se emplean datos de secuencia puede ocurrir la no independencia del estado del carácter y ello representar un problema.

En el caso de las secuencias de aminoácidos estas son generalmente codificadas de tal manera que cada residuo corresponde a un estado de carácter separado.

Sin embargo, frecuentemente un grupo de aminoácidos pueden considerarse como funcionalmente iguales bajo algunos criterios. Estos criterios y la agrupación de los aminoácidos particulares varia dependiendo del autor, pero usualmente están basados en la polaridad, estructura y otras características de los residuos.

3. ANTECEDENTES

En 1992 Brenner definió a la Biología molecular teórica como el retorno a una vieja tradición en la ciencia, una costumbre que consiste en la observación y la medida. De tal manera, el estudio de los genomas difiere tanto de la Biología experimental como la Astronomía difiere de la Física experimental, es decir se basan ambos en la observación: Genomimia sería un nombre apropiado para esta disciplina.

Para los Genomistas quienes miran el cielo genético, la invención de las técnicas de secuenciación pueden ser el equivalente para los astrónomos de la invención del telescopio.

En este contexto, y con el uso masivo de las PC (computadoras personales), junto a la continua expansión de secuencias de aminoácidos conocidas (estimadas hasta agosto de 1994 en; 36000 para PIR y 187000 para EMBL), además de su continuo incremento, y de la creación de diferentes bancos para conservar estas secuencias, se ha creado este tipo de investigación--

teórica a partir del análisis de estas secuencias. En cierto sentido la biología molecular esta basada fundamentalmente en las secuencias de las principales macromoléculas. Es así que el análisis de ellas por métodos de computación, constituye un intento por comprender fenómenos tan complejos como son las interacciones bioquímicas. Para cada proteína o polinucleótido, y algunas veces carbohidratos o lípidos se intenta proveer de un marco físico teórico de cómo esas secuencias interactuarían en el tiempo y espacio. De esta manera se ha pretendido unificar a la biología molecular teórica y a la experimental en las secuencias.

Dentro de los proyectos que se llevan a cabo en el Instituto de Biotecnología en la UNAM, se encuentra el de aislamiento de genes de las proteínas cpn60, por lo que con el fin de lograr la detección de estos por hibridación de ácidos nucleicos, se inicio una búsqueda de secuencias de estas proteínas y sus genes para utilizar las regiones más conservadas como sondas de hibridación de ácidos nucleicos.

Inicialmente se hizo una revisión preliminar en la literatura, dando como resultado, la aparente ausencia de zonas conservadas reportadas en estas proteínas. Sin embargo, utilizando diversos enfoques, logramos discernir dichas regiones conservadas, iniciándose de esta manera un análisis mas sistemático y detallado que condujeran a describir y caracterizar estas secuencias.

4 O B J E T I V O S

El presente trabajo, tiene como objetivos, el iniciar una búsqueda y analizar sistemática y detalladamente las secuencias de aminoácidos en la estructura primaria de las proteínas de la familia cpn60 (chaperoninas), las cuales se han reportado como altamente conservadas y de esta manera intentar discernir posibles implicaciones estructurales, funcionales y evolutivas. Para lograr estos objetivos se utilizaron métodos de computación.

El Proyecto estuvo a cargo de el M. en C. Víctor Valdés López en colaboración con el Biol. Antonio Lazcano-Araujo y el P. de B. Miguel Rangel Aguilera, este último como trabajo de tesis de licenciatura.

5 MATERIALES Y METODOS

Se utilizó el programa de computadora PC/Gene: Sistema software para análisis de secuencias de ácidos nucleicos y proteínas, versión 6.26, diseñado en la Universidad de Génova Italia.

Se utilizaron los siguientes bancos de genes:

(1) Protein Intelligentics Resource (PIR), de la Georgetown University Medical Center (National Biomedical Research Fundation). Con terminal instalada en el Instituto de Biotecnología de la UNAM.

(2) Genbank: Genetic Sequence Database Computer and Information Science Div. BBN Laboratories, Inc. Con terminal instalada en el laboratorio de Microbiología de la Facultad de Ciencias de la UNAM.

(3) EMBL Data library. Graham Cameron, Data Library Manager Postfach 10 2209 Meyerhofstrasse 1 6900 Heidelberg, Germany, con terminal instalada en el Instituto de Biotecnología de la UNAM.

Microcomputadora. AT 386 de 33 Mhz, con disco duro de 106 Megabytes, con memoria en Ram de 4 Megabytes, y monitor VGA.

Instalada en el laboratorio de Microbiología de la Facultad de Ciencias de la UNAM.

Programa de software elaborado por el Dr Luis Medrano (1993), para la identificación de las zonas conservadas.

Programa de software diseñado por el Dr Luis Medrano (1993), para ilustrar el promedio de identidad en un gráfica de matriz de puntos.

Programa de software SIGMA PLOT 5, instalado en el laboratorio de Microbiología de la Facultad de Ciencias de la UNAM.

Obtención de las secuencias.

Las secuencias obtenidas se guardaron en diskette para ser procesadas en el PC/Gene mediante el comando SEQUIN, el cual permite crear o modificar secuencias de ácidos nucleicos y proteínas. En este comando se le suprimió el péptido líder a las proteínas de eucariontes, y luego se utilizó la opción TRANSL, la cual tiene como función la traducción de las secuencias de nucleótidos a aminoácidos.

Secuencias obtenidas (incluyendo secuencia consenso), así como los nombres científicos de los organismos, las referencias bibliográficas, denominación de los bancos, tipo celular y el nombre que las secuencias tendrán en este trabajo.

Nombre científico	referencias	banco	secuencia
<i>Escherichia coli</i> proc(bacteria)	Hemmingsen <i>et al.</i> 1988	Genbank	Ecogroel
<i>Synechococcus sp</i> proc(cianobacter)	Webb <i>et al.</i> 1990	PIR	Groel\$synp7

<i>Synechocystis sp</i> proc(cianobacter)	Chitnis <i>et al.</i> 1991	EMBL	Groel\$syny3
<i>Micobacterium</i> <i>leprae</i> proc(micobacter)	Mhera <i>et al.</i> 1986	EMBL PIR	Micobovi
<i>Micobacterium</i> <i>tuberculosis</i> proc (eubact)	Shinnick <i>et al.</i> 1989	Genbank	Micotube
<i>Cricetulus</i> <i>griseus</i> euc(mitocon)	Picketts <i>et al.</i> 1989	EMBL PIR	P60\$ham
<i>Homo sapiens</i> euc(mitocon)	Waldinger <i>et al.</i> 1988	EMBL PIR	P60\$homo
<i>Mus musculus</i> euc(mitocon)	Venner y Gupta, 1990	EMBL PIR	P60\$mickey
<i>Rattus</i> <i>norvegicus</i> euc(mitocon)	Peralta <i>et al.</i> 1990	EMBL PIR	P60\$rata
<i>Sacharomyces</i> <i>cerevisiae</i> euc(mitocon)	Johnson <i>et al.</i> 1989	Genbank	P60\$yeast

<i>Arabidopsis thaliana</i> euc(cloropla)	Zabaleta <i>et al.</i> 1992	Genbank	Rubb\$ara sub beta
<i>Brassica napus</i> euc(cloropla)	Martel <i>et al.</i> 1990	EMBL PIR	Ruba\$brasi sub alfa
<i>Triticum aestivum</i> euc(cloropla)	Hemmingsen <i>et al.</i> 1988	EMBL PIR	Ruba\$trigo sub alfa
<i>Brassica napus</i> euc(cloropla)	Martel <i>et al.</i> 1990	EMBL PIR	Rubb\$brasi sub beta
<i>Sulfolobus shibatae</i> proc(arqueabacterias)	Trent <i>et al.</i> 1991	EMBL	TF55
Secuencia consenso	Este trabajo	---	Chapconsen

Alineamientos pareados de las secuencias, gráfica de matriz de puntos.

Este análisis se realizó con el objeto de comprobar la homología entre todas y cada una de las secuencias. Para ello se requirió del comando PALIGN de PC/Gene en el cual se utiliza el método desarrollado por Myers y Miller (1988), para el alineamiento de dos proteínas o nucleótidos.

Además de la comparación directa se utilizó un programa de computación diseñado por el Dr Luis Medrano (1993) de la Facultad de Ciencias de la UNAM, este presenta la identidad de dos secuencias en forma de matriz de puntos. Lo anterior se hace de la siguiente forma: se comparan dos secuencias una en el eje (Y) y otra en el eje (X), a cada residuo se le asigna un valor determinado, de esta manera cada valor que coincida en el espacio entre una y otra secuencia constituye un residuo conservado, es así que cuando coincidan varios residuos se construirá una recta de puntos, lo cual se interpretará como una zona de alta conservación entre esas secuencias. De esta manera se ilustra el grado de conservación.

Todas estas gráficas se construyen en el programa SIGMA PLOT en el cual se corren las matrices de puntos.

Con los resultados anteriores se construyeron las siguientes tablas; 1) distancias evolutivas; 2) índice porcentual de similitud (triángulo inferior) índice porcentual de identidad (triángulo superior); y 3) tabla de similitud por grupos filogenéticos.

En las tablas de similitud se cuantifican tanto los residuos conservados (invariantes), como los similares, y que pueden definirse, como el promedio general de conservación.

Para los objetivos anteriores se tomaron los siguientes criterios.

1) Las distancias evolutivas se calculan a partir de los resultados de los alineamientos pareados, para lo cual el total de los residuos hipervariables son multiplicados por 100 y dividido entre el número de la secuencia de menor tamaño (Doolittle, 1986).

2) Para el índice porcentual de similitud, y de identidad se toman los resultados directamente de las alineaciones pareadas (palign).

3) Los resultados de la tabla de similitud por grupos filogenéticos, se toman directamente de las alineaciones pareadas, donde el promedio de residuos invariables se suma al de los similares y se cuantifica el porcentaje.

Alineamientos múltiples de las secuencias.

En esta etapa se utilizó la opción CLUSTAL del Pcgene, que está basada en el método diseñado por Higgins y Sharp (1989), y que permite alinear hasta 25 secuencias de un máximo de tamaño de 1200 residuos o bases.

Derivación de una secuencia consenso.

Esta se llevó a cabo con base en el resultado de la alineación múltiple. Se prefirió hacerla en forma manual, utilizando los criterios de comparación de los rasgos de cada aminoácido, de acuerdo a normas definidas por Taylor (1986). La secuencia así obtenida se le denominó posteriormente Chapconsen, la cual en seguida se analizó en conjunto con todas las demás.

Gráfica de puntos (dot plots).

La información generada en los alineamientos múltiples permite identificar las regiones más conservadas de los polipéptidos. Los residuos idénticos y/o conservados a lo largo de la escala filogenética deben tener papeles estructurales y/o funcionales muy importantes. Para obtener esta información se usaron dos enfoques; en forma manual en papel milimétrico, y en un programa de computación diseñado por el Dr Luis Medrano (1993) de la Facultad de Ciencias de la UNAM.

Para la primera parte se utilizó papel milimétrico tabulado desde la posición 1 hasta la 550 de los residuos, en seguida se marcaron con puntos para cada posición de los residuos invariables (*) y de los similares (.). Como consecuencia de lo anterior, en la gráfica se advierten zonas de alta densidad en conservación así como regiones hipervariables. Con este resultado se trazó una gráfica de esos puntos en el cual cada punto corresponde a un 10% de conservación en un ventana de 10 posiciones. La segunda parte del método consiste en la construcción de un programa de computación cuyo algoritmo esta basado esencialmente en los mismos criterios, el resultado de este programa confirma el anterior resultado.

Además de la gráfica anterior se generó otra de variabilidad, basada esencialmente en los mismos criterios, y en el mismo programa.

Localización de agrupamientos o periodicidades en los aminoácidos.

Para esta fase se utilizó el comando PRESIDUE del PcGene. Este programa computa concentraciones locales de aminoácidos que sean de algún interés específico por ejemplo, si se decide cuantificar aminoácidos hidrofóbicos, el programa los detecta a través de la secuencia. Se observara entonces una gráfica con zonas ricas en concentración de estos residuos. La gráfica es producida por cálculos individuales para cada posición de la proteína. Las gráficas que aquí se presentan fueron realizadas en la secuencia consenso.

Análisis de perfiles hidropáticos y construcción de un perfil promedio.

En esta etapa se utilizó el programa SOAP de Pcgene, que gráfica índices de hidropatía de secuencias, de acuerdo al método de Kyte y Doolittle (1982). La hidropatía ha sido definida como la inclinación que tienen las cadenas de secuencias hacia la hidrofobicidad (o hidrofiliidad). El método trabaja asignando a cada residuo su valor hidropático, y posteriormente se suman y grafican esos valores.

Para construir un perfil promedio se utilizó el recurso de sobrelapar todas las gráficas individuales, resultando de ello una sola gráfica en la cual se identificaron las zonas promedio.

Análisis de diversos parámetros en la estructura secundaria.

Estos se llevaron a cabo con el Pcgene, y se utilizaron los métodos de Garnier *et al*,(1978);Gascuel y Goldmard, (1988); y de Novotny y Auffray,(1984), para la predicción de estructura secundaria. Estos tienen en promedio una validez de 56% a 57%, y están basados en estructuras terciarias que son conocidas, y se hacen de la siguiente manera; cada residuo requiere de una ecuación para cada uno de los estados conformacionales. La ecuación toma en cuenta la influencia de hasta 8 residuos antes y después del analizado.

Para obtener la estructura secundaria promedio, se empleó el criterio de utilizar la secuencia consenso, ya que se observó que conserva características generales de todas las secuencias, Además se usaron criterios basados en el trabajo de: Zvelebil *et al* (1987). todas estas predicciones funcionan solo para proteínas globulares.

Construcción de árboles filogenéticos.

Existen en general dos métodos para la construcción de árboles evolutivos. Uno de ellos es el llamado cladista, el cual se basa en análisis de las vías o rutas de los procesos evolutivos, que cuantifica cuantas ramas existe entre un grupo de organismos, cuantas y cuales se conectan entre sí, y cual es la secuencia evolutiva de estas ramas. A estos árboles se les denomina **cladogramas**. Así el cladograma es aquel árbol que muestra las relaciones de los descendientes por medio de un ancestro común.

El método más utilizado en este sistema es el de máxima parsimonia de Fitch, que refleja la naturaleza general de la dinámica de las mutaciones.

El principio de máxima parsimonia, involucra la búsqueda, y la identificación de un árbol filogenético, que requiera el menor número de cambios evolutivos ("mínima evolución") para explicar las diferencias observadas. A este árbol se le denomina árbol de máxima parsimonia. Este método es muy útil para construir relaciones filogenéticas a partir de datos fragmentados, y se usa cuando se quiere relacionar un número grande de especies (por ejemplo 60).

El otro método es el fenético. Por medio de este se analizan las relaciones entre un grupo de organismos en base a el grado de similitud entre ellos, esta similitud puede ser molecular, fenotípica o anatómica, el árbol que ilustra relaciones fenéticas es llamado fenograma, en este sistema los datos de las secuencias son convertidos a medidas de distancias, posteriormente estas medidas son utilizadas para la construcción de árboles.

En este trabajo, el método que se empleó para analizar las relaciones filogenéticas de cpn60 fue el fenético, debido a que las secuencias, aunque completas son aun escasas (14 en el momento de iniciar este trabajo). En este sentido, la construcción del árbol filogenético, involucra a la segunda etapa del comando CLUSTAL del Pcgene para alineamientos múltiples, que a su vez esta basado en las similitudes obtenidas en la primera etapa del método, este es referido como UPGMA ("Unweighted Pair Group Method with Arithmetic mean") también conocido como análisis promedio de agrupamiento. Este método reúne inicialmente las dos secuencias más similares, posteriormente la siguiente más parecida y así sucesivamente.

6 RESULTADOS

Alineamientos pareados.

Se llevaron a cabo los alineamientos pareados de todas las secuencias disponibles, si bien en este trabajo solo se muestran tres ejemplos de estas alineaciones (en la figura 1f, se pueden consultar todos los resultados). La figura 1a, muestra el primer alineamiento donde se comparó Rubb\$bra (subunidad beta de *Brassica napus*) vs P60\$yeas (*Sacharomyces cerevisiae*), la figura 1b muestra el alineamiento de GroL\$sy7 (*Synechococcus sp*) vs P60\$rata (*Rattus norvegicus*). Como se puede observar estos alineamientos no muestran diferencia significativa a lo reportados en la literatura (Hemingsen *et al.* 1988;Pickets *et al.* 1989;Venner y Gupta,1990).

Si bien es claro el alto grado de identidad de esta proteína, a simple vista no se observan claramente regiones de conservación, por lo que para descubrir estas posibles zonas se construyeron gráficas de matriz de puntos. Como ejemplo de lo anterior en la figura 1d, se pueden observar probables zonas de conservación en P60\$yeast (*Sacharomyces cerevisiae*) vs Rubb\$arat (subunidad b de *Arabidopsis thaliana*), cuyo promedio de identidad es 44.9 % (ver figura 1f), otro ejemplo se muestra en la figura 1e donde se ilustran posibles zonas de conservación en Ecogroel (*Escherichia coli*) vs Micobovi (*Micobacterium leprae*), con un promedio de conservación 58.7%.

La figura 1f muestra que los índices de identidad entre todas las cadenas de cpn60, se encuentran en el rango de 40-50%.

Tomando en cuenta el criterio de Doolittle (1986), de que a partir de una identidad mínima de 20% entre dos secuencias estas se consideran homólogas, entonces consideramos que todas las secuencias de cpn60 aquí descritas son homólogas. Además, se puede considerar que esta proteína ha estado muy conservada a lo largo de la evolución, ya que los índices de identidad obtenidos (40 - 50 %) son muy superiores al valor de decisión (20%). Por lo tanto, el reloj molecular de cpn60 debe ser muy lento.

Estos resultados son corroborados por los descritos en la figura 1g, donde se observan las distancias evolutivas, y en las cuales se obtienen valores de 20 a 30 %, mismas que son muy bajas, e indican una estrecha relación filogenética.

Con el fin de evidenciar las relaciones filogenéticas de los diferentes grupos representados, se muestra en la figura 1h los valores de identidad de cpn60 organizados por grupos filogenéticos, en la figura 1h (a) se puede observar que cianobacterias y las demás eubacterias presentan entre si el mayor grado de similitud de entre todos los grupos (68.3%-73.4%). En la figura 1h (b) se observa que al comparar las cpn60 alfa y beta de planta estas no muestran un grado mayor de similitud (67.9%-69.9%). En este sentido, Musgrove (1989), demuestra que ambas subunidades son codificadas por dos genes diferentes.

Así cuando se comparan estas secuencias con cianobacterias y eubacterias (fig 1h (c)) se aprecia que cloroplastos de plantas y cianobacterias son altamente similares (69.5%-73.8%). Por lo contrario cuando las secuencias de planta se comparan con las secuencias de mitocondria (fig 1h (d)) el valor de similitud obtenido es claramente menor (62.7%-65.6%), en esta misma comparación se puede observar que a su vez las secuencias de mitocondria son más parecidas a las de la eubacteria *Escherichia coli* (67.5%-69.5%), que a las de cianobacterias (66.5%-67.6%), y micobacterias (64.4%-65.0%).

Los resultados anteriores sugieren, que si bien entre todas las cpn60 existe un alto grado de conservación, consistente en una función similar entre ellas, no obstante en organismos y ambientes específicos existen presiones de selección que afectan la estructura y la función de estos polipéptidos. Así es posible suponer que los requerimientos para el plegamiento de proteínas en sistemas relacionados con la fotosíntesis (planta/cianobacteria), sean ligeramente diferentes a los de eubacteria/mitocondria. Dicho de otra manera, si bien la función global de las chaperoninas requiere una estructura general, la variedad de medio ambientes, y de sustratos (polipéptidos), imponen requisitos especiales y específicos que se manifiestan en diferencias y/o similitudes a nivel de estructura primaria de las cpn60.

El caso de las Arqueobacterias.

En el caso de arqueobacterias, ha sido reportada una proteína relacionada a cpn60 en *Sulfolobus shibatae* (Trent *et al.*1991), y ha sido llamada TF55, la cual fue analizada conjuntamente con la cpn60.

Sin embargo, como se muestra en la fig 1c, su patrón de identidad es muy difuso. Como se puede observar en la fig 1h, la proteína resultó altamente divergente en secuencia de aminoácidos; su promedio de identidad con cpn60 va desde 14.0% con levadura, hasta 19.1% con micobacterias, es decir que esta abajo de el promedio de decisión, en una zona que describe Doolittle (1986), como región de penumbra o de duda, es así que dado su nivel bajo de identidad de TF55 con las chaperoninas de eucariontes y eubacterias, se sugeriría que su divergencia de las otras proteínas fue muy temprana, pero por otra parte con este promedio de identidad, también se dudaría incluirla dentro de la familia de las cpn60. Como ya se dijo antes, Doolittle (1986), propone un mínimo 20% de identidad en las secuencias de aminoácidos para decidir que dos proteínas son homólogas, esto es en base a una serie de pruebas que se hicieron a diversas secuencias, las cuales se sabía que no tenían ninguna relación, y que cuando eran alineadas presentaban hasta 15% de identidad, no obstante el promedio general de identidad debida al azar es 5 a 6% (sin gaps), es por esto que en el caso de las cpn60 de arqueobacterias, dado su bajo promedio de identidad, se sitúan en un limite de penumbra, en el cual pueden ser o no homóloga a cpn60. Es por lo anterior que se ha tratado de relacionarla a cpn60 por otros medios (estructura secundaria, hidropatía, dominios etc).

De lo anterior solo el análisis de la proteína por dominios resultó positivo, encontrándose en el alineamiento múltiple con cpn60 dos regiones compatibles con esta proteína, el primer dominio de TF55, es contraparte de la primer zona de alta similitud en cpn60, se ubica del residuo 38 al 137, y tiene un nivel de identidad aproximado de 20%, con un promedio de similitud cercano al 33% (fig 2a y 2b).

El segundo dominio corresponde a la cuarta zona de máxima identidad en cpn60, destacándose este hecho ya que quizás esta zona sea la de mayor importancia en cpn60. Este motivo está entre el residuo número 401 y el 450, y presenta un promedio de identidad de 22%, con 34% de similitud (fig 2a y 2b). Además de lo anterior, TF55 conserva las funciones esenciales de molécula chaperona, es decir, presenta actividad de choque térmico, tiene una estructura cuaternaria parecida a las chaperoninas (Phipps *et al.*1993), y se cree que su mecanismo de acción es similar, ya que como cpn60, ella posee actividad débil de ATPasa (Saibil y wood, 1993,).

En vista de las diferencias y las similitudes encontradas en esta proteína para las arqueobacterias en relación a las bacterias y los eucariontes, es difícil decidir por criterios tradicionales si cpn60 es o no homóloga para los tres linajes. Ello se puede deber a que las arqueobacterias tuvieron también el gen ancestral, pero que este ha divergido en forma muy diferente a sus contrapartes eucariontes y bacterias, por razones que se desconocen.

Otra razón sería que cpn60 de arquea, no sea un gen homólogo, y en consecuencia estaríamos hablando de otra proteína, tan antigua como cpn60, pero diferente a ella, en cuyo caso se trataría de una forma de evolución convergente. Nosotros creemos que se trata de el primer caso, basándonos principalmente en el análisis de los dominios, cuyo nivel de identidad (20% y 22%) son suficientes a el criterio establecido, es decir que cpn60 es homóloga desde este punto de vista.

Fig 1a).Ejemplo de alineamiento pareado. Cloroplasto sub β *Brassica napus*. vs mitocondria levadura *Sacharomyces cerevisiae*. Los caracteres invariables se pueden distinguir con un asterisco (*), los caracteres similares se distinguen con un punto (.), las diagonales (-) representan huecos, los cuales se utilizan para optimizar la alineación. El total de residuos para levadura es de 550, y para Brassica es de 534. El número de residuos invariables (*) es 243, lo que representa el 45.5% del total de la alineación; el número de residuos similares (.) es de 102, representando 19.1%; la suma de residuos invariables + similares es 345, refiriendo el 64.6% de conservación total (ver fig 1f).

```

RUBB$BRA   AAKELHFNKDGT TIRKLOAGV NKLADLVGV T L GPKGRNVV LESKYGSPRIVNDGV
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   SHKELKFGVEGRA--SLLKGVETLAEAVAATL GPKGRNVLI EQFPFPFKITKDGV

RUBB$BRA   TVAREVELED PVENIGAKLVRQAAAKTNDLAGDGT TTSV VLAOGFIAEGVKV VAA
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   TVAKSIVLKD K FENMGAKLLQEVASKTNEAAGDGT TSATV LGRAIFTESVKNVAA

RUBB$BRA   GANPVLITR GIEKTAKALVAELKMSKEVEDSE-LADVA AVSAGNNAEIGSMIAE
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   GCNPMDLRRGSQVAVEK VIEFLSANKKEITTSEEIAQVATISANGDSHV GKKLLAS

RUBB$BRA   AMSRVGRKG VV TLEEGKSAENALYVVEGMQFDRGYVSPYFVT DSEKMSVEFDNCK
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   AMEKV GKEGVITIREGR TLEDELEVTEGMRFRDRGFISPYFITDPKSSKVEFEKPL

RUBB$BRA   LLLVDKKITNARDLVG VLEDAIRGGYPILIIAEDIEQEALATLVVNKLRGTLKIA
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   LLLSEKKISSIQDILPALEISNQSR RPLIIAEDVDGEALAACILNKL RGVKVC

RUBB$BRA   ALKAPGFGERK SQYLDDIAILT GATVIREEVGLSLDKAGKEVLGHAAKV VLTKET
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   AVKAPGFGDNRKNTIGDIAVLTGGTVFTEELDLKPEQCTIENLGSCDSITVTKED

RUBB$BRA   STIVGDGSTQDAVOKRV TQIKN LIEQAEQD-YEKEKLN ER IAKLSGGVAVIQVGA
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   TVILN GSGPKEA IQERIEQIKGSIDIT TNSYEKEK LQERLAKLSGGVAVIRVGG

RUBB$BRA   QTETELKEK KLRVEDALNATKAAVEEGIVVGGGCTLLRLASKVD A IKAALDNDEE
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   ASEVEVGEKKDRYDDALNATRAAVEEGILP GGGTALVKASRVLDE--VVVDNFDQ

RUBB$BRA   KVGADIVKRALS YPLKLI AKNAGVNGSVVSEKVL SN--DNV KFGYNAATGKYEDL
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   KLGVDIIRKAITRPAKQIIENAGEEGSVIIGKLIDEY GDDFAKGYDASKSEY TDM

RUBB$BRA   MAAGIIDPTKVVRC CLEHAASVAKTFLMSDCVVVEIKEPEPV-----
          . * * * * * . * * * * * . * * * * * . * * * * * . * * * * *
P60$YEAS   LATGIIDPFKVVRSGLVDASGVASLLATTEVAIVDAPEPPAAAGAGGMPGMPGM

RUBB$BRASI- ---P -534 60$YEAST - PGMM -550

```


Fig 1c. Ejemplo 3. Archeobacteria *Sulfolobus shibatae* (tf55) vs Cloroplasto subunidad β de *Brassica napus*. Los asteriscos (*) refieren residuos invariables, los puntos (.) ilustran los aminoácidos conservados, las diagonales (--) interpretan huecos, necesarios para mejorar la alineación. El total de residuos para brassica es 534, para Sulfolobus es 552. El porcentaje de residuos invariable es 16.5%, lo que representa 88 aminoácidos, el promedio de similaridad es 15.7%, representando 84 residuos, la conservación total es la suma de residuos invariables + similares, dando un total de 172 (32.2%) (ver fig 1f).

```

TF55      MATATVATTPEGIPVIJLKEGSSRTYGKEALRANIAAAVKAIEEALKSTYGPRGMD
          *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  AAKE-LHFNKDGTTIRKLQAGVNKL---ADLVGVTLGPKGRNVVLESKYGSPR--

TF55      KMFVDSLGDITITNDGATILDKMDLQHPTGKLLVQIAKQODEETADGKTAVILA
          *   *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  --IVN--DGVTVAREVE--LEDP--VENIGAKLVRQAAAKTNDLAGDGTTTSSVLA

TF55      GELAKKAEDLLYKEIHPTIIVSGYKKAEEIALKTIODIAQPVSIINDTDVLRKVAL
          *   *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  QGFIAEGVKVVAAGANPVLITRGI EKTAKALVAELKKMSKEVE--DSELADVAAV

TF55      TSLGSKAVAGAR-EYLADLVVKAQAQVAELRG-DKWYVDLDNVQIVKHHGGSIND
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  SAGNNAEIGSMIAEAMSRVGRKGVVTL EEGKSAENALYVVEGMQFDRGYVSPYFV

TF55      TQLVYGIVVDKEVVHPCMPKRIENAK--IALLDASLEVEKPELDAEIRINDPTOM
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  TDSEKMSVEFDNCKLLLVDKKITNARDLVGVLEDAIRGGYPILIIAEDIEQEALA

TF55      HKFLEEEENILKEKVDKIAATGANVVICQKGI DEVAQHYLAKKGI-LAVRRAKKS
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  TLVVNKL RGT LKIAALKAPGFGERKSQYLD DIIAILT GATVIREEVGLSLDKAGKE

TF55      DLEKLARATGGRVISNI-DELTSODLGYAALVEERKVGED-KMVFVEGAKNPKSV
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  VLGHAAKVVLTKETSTIVGDGSTQDAVQKRVT OIKNLI EQAEQDYEKEKLNERIA

TF55      SILIRGGLERVVDETERALRDALGTVADVIRDGRAVAGGGAVEIEIAKRLRKYAP
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  KLSGGVAVIQVGAQTETELKEKLRVEDALNATKAAVEEGIVVGGGCTLLR----

TF55      QVGGKEQLAIEAYANAIEGLIMILAENAGLDPIDKLMQLRSLHENETNKWYGLNL
          *   *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  -LASKVDAIKAALDNDEEKVGADIVKRALSYP LKLI AKNAGVNGSVVSEKV---L

TF55      FTGNPEDMWKLGVIEPALVKMNAIKAATEAVTLVLRIDDIVAAGKKGSGEPGGKK
          *   *   *   *   *   *   *   *   *   *   *   *
RUBB$BRA  SNDNVKFGYNAATGKYEDLMAAGIIDPTKVVR CCL--EHAASVAKTFLMSDCVVV

TF55      - EKEEKSSSED -552
          *   *
RUBB$BRASI- EIKEPEPVP -534

```

fig 1d). Comparación pareada de secuencias en matriz de puntos. En esta gráfica se distinguen posibles zonas de identidad en dos secuencias de cpn60 (eje Y de mitocondria P60\$yeast *Sacharomyces cerevisiae*, su número de residuos es de 550. Eje X de cloroplasto Rubb\$arat *Arabidopsis thaliana*, con número de residuos de 546). Cuyo promedio de identidad (residuos invariables) es 44.9% (ver fig 1f).

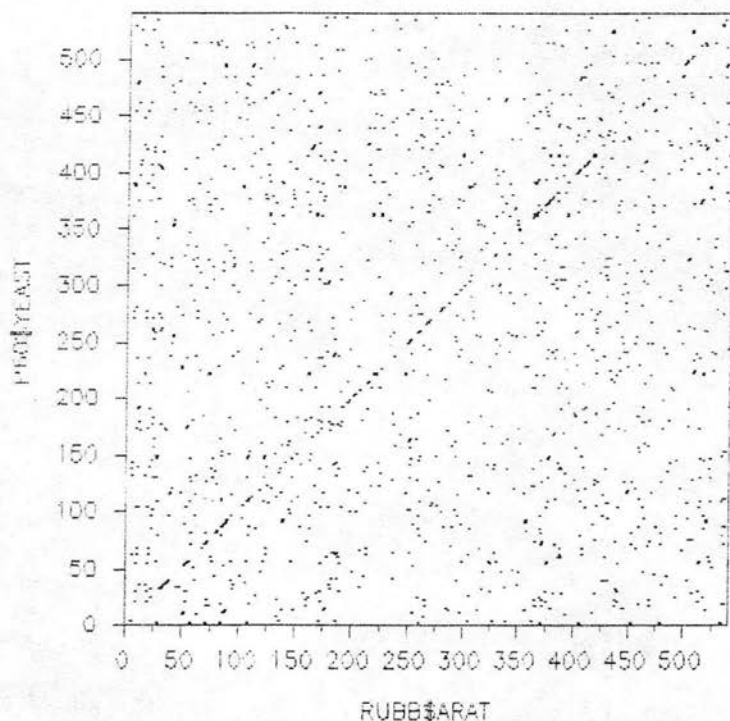


fig 1e). Comparación pareada de identidad en matriz de puntos, en esta gráfica, como en la anterior, se ilustra el promedio de identidad de una alineación pareada (eje Y Ecogroel *Escherichia coli*, con número de residuos de 548. Eje X Micobovi *Micobacterium leprae* su número de residuos es 540). Con promedio de identidad de 58.7% (ver fig 1f).

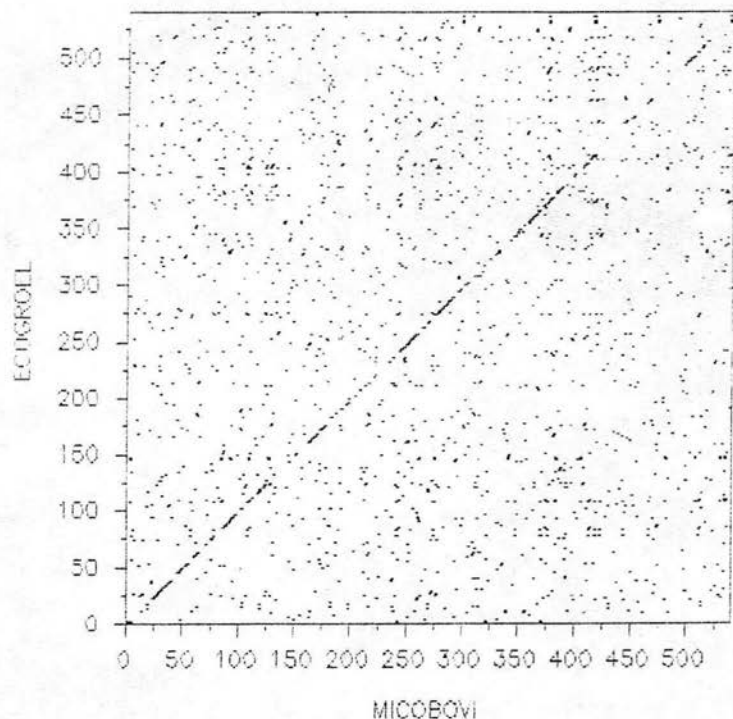


fig 1f) Índice porcentual de identidad (triángulo superior, solo residuos invariables), **índice porcentual de similitud** (triángulo inferior, residuos similares + invariables), esta figura fue construida con base en los resultados de los alineamientos pareados, como se puede apreciar los niveles de identidad están en promedio de 40% a 50%, por lo que están por encima de el nivel de decisión (20%), en consecuencia todas las secuencias son homologas (la nomenclatura de las secuencias están en material y métodos).

	ECOGR	CHCON	MICOB	SYNP7	SYNY3	MICOT	BARAT	BBRAS	ABRAS	ATRIG	HANST	MICK	RATA	HOMO	YEAS
ECOGROEL	---	63.8	58.7	55.3	55.8	52.7	51.5	52.2	47.8	47.8	50.2	49.6	50.8	49.8	54.8
CHAPCONSEN	73.6	---	65.5	64.2	66.8	57.7	63.6	64.8	63.8	61.4	54.9	54.9	55.1	49.8	56.6
MICOBV1	72.8	73.7	---	59.6	56.9	68.7	52.6	52.8	51.1	48.7	47.8	47.8	48.8	47.8	47.8
GR0L\$SYNP7	71.8	75.1	72.9	---	62.5	51.8	53.5	55.4	53.9	51.9	45.2	45.8	45.4	45.8	49.1
GR0L\$SYNY3	73.3	77.5	71.5	75.9	---	53.2	56.8	56.6	57.4	55.6	46.5	46.2	46.5	46.2	48.7
MICOTUBE	67.2	59.2	73.5	68.3	69.2	---	46.4	46.4	48.3	45.1	41.6	41.4	41.7	41.4	43.6
RUBB\$ARATH	69.1	73.6	67.8	78.8	71.8	63.8	---	97.2	58.6	58.5	46.7	46.3	46.8	46.2	44.9
RUBB\$BRASI	69.2	74.1	68.8	78.4	72.8	63.8	98.7	---	58.8	58.8	47.2	47.2	46.8	46.8	45.5
RUBA\$BRASI	66.8	73.6	67.1	78.8	73.8	64.7	69.8	69.9	---	81.6	44.6	44.1	43.9	44.1	45.9
RUBA\$TRIG0	66.2	73.5	66.3	69.5	73.3	63.8	67.9	69.2	98.7	---	42.9	42.5	42.7	42.3	43.1
P68HAN	67.5	68.4	64.7	64.7	66.6	61.1	64.5	65.8	63.2	61.2	---	98.7	98.7	97.1	56.8
P68MICKY	67.3	68.3	64.5	64.5	66.5	68.9	64.2	64.8	63.1	61.8	99.6	---	99.6	97.6	56.2
P68RATA	67.7	68.5	64.9	64.7	66.7	61.2	63.8	64.2	62.7	61.2	99.6	99.6	---	97.6	56.8
P68HOMO	67.5	68.2	64.4	65.8	66.5	61.9	64.5	65.8	63.1	68.8	98.9	98.5	98.5	---	56.4
P68YEAST	69.5	78.8	65.8	66.9	67.6	62.8	63.4	64.6	65.6	63.1	72.8	72.6	72.6	73.8	---

fig 1g). Tabla de distancias evolutivas. Esta figura fue construida con base en las alineaciones pareadas, en estos resultados se sustituyeron los residuos hipervariables, cuantificando esto como distancias evolutivas o como tasa de cambio de las diferentes especies, y como se puede observar estos arrojan un promedio de 20% a 30%, cambio que se considera bajo, es decir que esta proteína ha cambiado muy poco (ver nomenclatura de secuencias en material y métodos).

	ECOSRO	CHAPCON	MICOB0V	SYNP7	SYNY3	MICOTU	BARATH	BBRASI	ABRASI	ATRIGO	HAM	MICKY	RATA	HOMO	YEAST
ECOSROEL															
CHAPCONS	26.34														
MICOB0VI	27.96	26.34													
GR0L\$SY7	28.12	24.86	27.03												
GR0L\$SY3	26.64	22.44	28.51	24.00											
MICOTUBE	32.83	30.79	26.53	31.72	30.24										
RUBB\$ARA	30.95	26.34	32.96	29.96	28.93	36.17									
RUBB\$BRA	30.71	25.84	33.02	29.58	28.08	36.14	01.31								
RUBA\$BRA	33.27	26.39	32.89	29.18	26.28	35.31	31.84	30.14							
RUBA\$TRI	33.82	26.53	33.70	30.49	26.61	36.17	32.16	30.89	09.29						
P60HAM	31.75	31.53	35.37	35.29	33.39	38.96	35.53	35.01	36.00	38.81					
P60MICKY	32.66	31.72	35.55	35.47	33.57	39.14	35.71	35.20	36.98	39.00	0.36				
P60RATA	32.29	31.53	35.18	35.29	33.75	38.77	36.26	35.76	37.36	38.81	0.36	0.36			
P60HOMO	32.48	31.72	35.55	34.92	33.57	39.14	35.53	35.01	36.98	39.00	1.09	1.45	1.45		
P60YEAST	30.47	30.85	35.00	33.00	32.36	37.29	36.63	35.39	34.38	36.96	27.18	27.37	27.63	27.63	

1h) Tabla de similitud por grupos filogenéticos;

a) bacterias vs cianobacterias: b) subunidad alfa de cloroplasto vs subunidad beta de cloroplasto: c) subunidad alfa y beta de cloroplasto vs cianobacterias; bacterias y micobacterias: d) mitocondria vs bacterias; cianobacterias; subunidad alfa y beta de cloroplasto, y micobacterias: e) arqueabacteria vs subunidad alfa y beta de cloroplasto; mitocondria; bacteria; cianobacteria y micobacteria. En el inciso e, además del promedio de similitud también se incluyó el promedio de identidad, esto con el fin de evidenciar el bajo índice de arqueabacteria en relación a los demás grupos filogenéticos.

a)

	Synp7	Syny3
Ecogroel	71.8	73.3
Micobovi	72.9	71.5
Micotube	68.3	69.2

b)

	Ruba\$bras	Ruba\$tri
Rubb\$arath	69.0	67.9
Rubb\$brasi	69.9	69.2

c)

	Syn7	Syny3	Ecogr	Micot
Rubb\$arat	70.0	71.0	69.1	63.8
Rubb\$bras	70.4	72.0	69.2	63.8
Ruba\$bras	70.8	73.8	66.8	64.7
Ruba\$trig	69.5	73.3	66.2	63.8

d)

	Ecog	Syny3	Rubb	Ruba	Micb
P60\$ham	67.5	66.6	64.5	63.1	64.7
P60\$mick	67.3	66.5	64.2	63.1	64.5
P60\$rata	67.7	66.7	63.8	62.7	64.9
P60\$homo	67.5	66.5	64.5	63.1	64.4
P60\$yeas	69.5	67.6	64.6	65.6	65.0

e)

	similaridad	TF55 identidad
Rubb\$bra	32.2	16.5
Ruba\$bra	31.4	17.3
P60\$homo	31.2	16.2
P60\$yeas	25.8	14.0
Grol\$sy3	35.3	19.1
Ecogroel	32.2	16.1
Micobovi	32.2	19.1

Alineamientos múltiples.

En la fig. 2a se presentan los resultados de las alineaciones múltiples de cpn60 a lo largo de la escala filogenética (no incluye arqueobacterias), los caracteres invariables se distinguen con asteriscos (*), los conservados en puntos (.), en negritas las regiones que se han definido como de alta conservación. Las secuencias alineadas constan en promedio de 544 residuos, de estos 95 (16.9%) son invariables, 195 (34.7%) son similares. Así, el 51.6% de los aminoácidos (similares + invariables) son conservados. De ellos sobresale el bajo número de cisteínas (.45%) en contraste a el promedio global (3.2%) de otras proteínas (Dayhoff, 1978) (datos no mostrados). Ello sugiere que su ausencia no es determinante en la estabilidad estructural de las cpn60, por lo que quizás estas características estén dadas por otros factores, como podrían ser núcleos hidrófobos. Por otra parte la proteína es abundante en; ala (11.8%), glu (8.55%) y asp (6.21%), y particularmente escasa en; his (0.56%) y arg (3.90%) (datos no mostrados), lo que le confiere un carácter ácido. Además de lo anterior, en este alineamiento se puede apreciar que cpn60 es baja en prolinas (2.99%) en contraste al promedio global de otras proteínas (5.5%) (datos no mostrados). Sin embargo es importante hacer notar que varias de las posiciones invariantes en el alineamiento son prolinas. Por ello es posible que este residuo podría ejercer una importante función en la desestabilización de la estructura secundaria para permitir plegamientos específicos. En la figura 2b, se observa en negritas las regiones que se han definido como homologas de este péptido de arquea vs cpn60 de eucaria y eubacteria, se puede distinguir el escaso nivel de identidad que exhiben ambas proteínas. Como ya se menciono antes su grado de identidad se sitúa en nivel de penumbra (ver el caso de arqueobacteria).

ECOGROEL
MICOBOVI
MICOTUBE
GROL\$SY7
GROL\$SY3
RUBB\$ARA
RUBB\$BRA
RUBA\$BRA
RUBA\$TRI
P60\$HAM
P60\$MICK
P60\$RATA
P60\$HOMO
P60\$YEAS

DTTTIIDGVGE--EAAIOGRVAGIROQIEEATSD-YDREKLOERVAKLAGGVAVI
DETTTVEGAGD--TDAIAGRVAQIRQEIENS DSD-YDREKLOERLAKLAGGVAVI
DDTVIVDGGGT--AEAVANRAKHLRAEIDKSDSD-WDREKLGERLAKLAGGVAVI
DNNTTIVAEGNE---AAVKARVDOIRROIEETESS-YDKEKLOERLAKLSGGVAVV
DNTILVAGAHKRRSAGVKERIEQLRKEYAASDSD-YDKEKIQERIAKLAGGVAVI
ETSTIVGDGSGTO--DAVKRVRTQIKNLEQAEQD-YEKEKLNERIAKLSGGVAVI
RUBB\$ARA
RUBB\$BRA
RUBA\$BRA
RUBA\$TRI
P60\$HAM
P60\$MICK
P60\$RATA
P60\$HOMO
P60\$YEAS

... * ** .. ** .. ** .. ****

ECOGROEL
MICOBOVI
MICOTUBE
GROL\$SY7
GROL\$SY3
RUBB\$ARA
RUBB\$BRA
RUBA\$BRA
RUBA\$TRI
P60\$HAM
P60\$MICK
P60\$RATA
P60\$HOMO
P60\$YEAS

KVGAATEVEMKEKKARVEDALHATRAAVEEGVAVAGGVALIRVASK-LADLRGK-
KAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLLOAAP-TLDELK--L
KVGAATETALKERKESVEDAVAAKAAVEEGIVPGGGASLIHQARKALTELRASL
KVGAATETEMKDRKLRLLEDAINATKAAVEEGIVPGGGTTLAHLAPOLEEWATANL
KVGAATETELKDRKLRIEDALNATKAAVEEGIVPGGGTTLIRLAKGIESFKAQLS
QVGAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCTLLRLASKVDAIKATLD
QVGAQTETELKEKKLRVEDALNATKAAVEEGIVVGGGCTLLRLASKVDAIKAALD
KVGAATELEDRKLRIEDAKNATFAAIEEGIVPGGGATLVHLSTVIPAIKEKLE
KVGATTETELEDRLRIEDAKNATFAAIEEGIVPGGGAAYVHLSTYVPAIKETIE
KVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPALDSLKPS--
KVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPALDSLKPA--
KVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGGCALLRCIPALDSLTPA--
RVGGASEVEVGEKKDRYDDALNATRAAVEEGILPGGGTALVKASRVLDEVVVD--

.. * .. ** .. ** .. ** .. ** ..

ECOGROEL
MICOBOVI
MICOTUBE
GROL\$SY7
GROL\$SY3
RUBB\$ARA
RUBB\$BRA
RUBA\$BRA
RUBA\$TRI
P60\$HAM
P60\$MICK
P60\$RATA
P60\$HOMO
P60\$YEAS

NEDQNVGKIVALRAMEAPLROQIVLNCGEEPSVVANTV-KGGDGNV--GYNAATEE
EGDEATGANIVKVALEAPLQKIAFNGLPEFVVAEKV-RNLPAGH--GLNAQTV
TGDEVLGVDVFSEALAAPLFWIAANAGLDGSSVVVKKV-SELPAGH--GLNVGT
SGEELTGAQIVARALTRALKRIAENAGLNGAVISERV-KELPFDE--GYDASNQ
NDEERVAADI IAKALEAPLHQLASNAGVEGSLIVEKV-KEATGNQ--GYNVITGK
NDEEKVGADIVKRALSYPKLI AKNAGVNGSVVSEKVLSDNDNVKF--GYNAATGK
NDEEKVGADIVKRALSYPKLI AKNAGVNGSVVSEKVLSDNDNVKF--GYNAATGK
DADERLGDADIVQKALVAR-SLIAQNAIGIGEVVVEKI-MFSEWEI--GYNAMTDT
DHDERLGADIIQKALQAPASLIANNAGVEGEVVEKI-KESEWEM--GYNAMTDK
NEDQKIGIEI IKRALKIPAMTIAKNAGVEGSLIVEKILQSSSEI---GYDAMLGD
NEDQKIGIEI IKRALKIPAMTIAKNAGVEGSLIVEKILQSSSEV---GYDAMLGD
NEDQKIGIEI IKRALKIPAMTIAKNAGVEGSLIVEKILQSSSEV---GYDAMLGD
NEDQKIGIEI IKRALKIPAMTIAKNAGVEGSLIVEKIMQSSSEV---GYDAMAGD
NFDQKLGVDIIRKAITRPAKQI IENAGEEGSVI ICKLIDEGYDDFAKYDASKSE

... * ** * ** ..

```

ECOGROEL      YGNMIDMGILDPTKVTRSALQYAASVAGLMIITTECMVTDLPKNDAAADLGAAGGMG
MICOBOVI      YEDLLAAGVADPVKVVTRSALQNAASIAGLFLTTEAVVADKPEKEKASVPGGGDMG
MICOTUBE      YGDLAADGVIDPVKVVTRSALVNASSVARMVLTTETVVVDKPAKAE----DHDHHH
GROL$SY7      FVNMFTAGIVDPAKVTRSALQNAASIAAMVLTTETECIVVDKPEPK-----EKAPAG
GROL$SY3      IEDLIAAGIIDPAKVVRSAALQNAASIAAGMVLTTALVVEKPEPAPPAMPDMGGMG
RUBB$ARA      YEDLMAAGIIDPTKVVRCCLEHAASVAKTFLMSDCVVVEIKEPEPVPVGNPMDNS
RUBB$BRA      YEDLMAAGIIDPTKVVRCCLEHAASVAKTFLMSDCVVVEIKEPEPVP-----
RUBA$BRA      YENLLEAGVIDPAKVTRCALQNAASVAGMVLTTQAIIVVDKPKPKAPTAAPPQG-L
RUBA$TRI      YENLIESGVIDPAKVTRCALQNAASVSGMVLTTQAIIVVEKPKPKPKVAEPAEGQL
P60$HAM       FVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDPGMGAMGGM
P60$MICK      FVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEAVVTEIPKEEKDPGMGAMGGM
P60$RATA      FVNMVEKGIIDPTKVVRTALLDAAGVAPLLTTAEAVVTEIPKEEKDPGMGAMGGM
P60$HOMO      FVNMVEKGIIDPTKVVRTALLDAAGVASLLTTAEVVVTEIPKEEKDPGMGAMGGM
P60$YEAS      YTDMLATGIIDPFKVVRSGLVDASGVASLLATTEVAIVDAPEP---PAAAGAGGM

```

.. . * . ** ** * . *

ECOGROEL	GMGGMGGM---M	548
MICOBOVI	GMD-----F	540
MICOTUBE	GHA-----H	539
GROL\$SY7	AGGGMGDF-DY-	544
GROL\$SY3	GMGGMGGM-GMM	551
RUBB\$ARA	GLG-----Y	546
RUBB\$BRASI	-----	534
RUBA\$BRASI	MV-----	538
RUBA\$TRIGO	SV-----	541
P60\$HAM	G-GMGG--GMF	548
P60\$MICKEY	G-GMGG--GMF	548
P60\$RATA	G-GMGG--GMF	548
P60\$HOMO	G-GMGG--GMF	548
P60\$YEAST	P-GGMPGMPGMM	550

Número de
residuos

Identidad	*	95	(16.9%)
Similitud	.	195	(34.7%)
Promedio global		290	(51.6%)

Es notable que el promedio general de conservación resulte tan alto, sobre todo una vez que se han alineado hasta 14 secuencias de amplio rango filogenético, ya que solo un grupo (arqueobacteria) esta ausente en esta alineación, estos resultados enfatizan sin duda la amplia relación filogénica de esta proteína, y a la vez sugieren un origen común para este péptido en los diferentes grupos.

Figura 2b alineamiento múltiple 2: Esta figura muestra el alineamiento de cpn60 en grupos filogenéticos que representan a los tres linajes primarios universales (Arqueobacterias, Eubacterias y Eucariontes). Las regiones que probablemente son homólogas se dividen en dos dominios, mismos que estarían en las siguientes posiciones; primer dominio residuo 38-137, (nivel de identidad 20%, similitud 33% dato no mostrado): El segundo dominio estaría en el residuo 401-450, (identidad 22%, similitud 34% dato no mostrado), estos dominios se pueden distinguir en negritas. Los residuos invariables están en asteriscos (*), los similares se notan con puntos (.), las rayas(-) representan huecos, los cuales son utilizados para optimizar la alineación, el número total de residuos de las secuencias se encuentran al final, así como el promedio de identidad, el de similitud, y el promedio general de conservación. La nomenclatura de las secuencias está en material y métodos.

38

TF55	MATATVATTPEGIPV I I LKEGSSRTYGKEALRANIAAVKAIEEALKSTY G
ECOGROEL	MAAKDVKFGNDAR--VKMLRG-----VNVLADAVKVTLG
MICOBOVI	MA-KTIAYDEEAR--RGLERG-----LNLADAVKVTLG
GROL\$SY3	MS-KLISFKDESR--RSLEAG-----INALADAVRI TLG
RUBB\$BRA	-AAKELHFNKDGTTIRKLQAG-----VNKLADLVGVTLG
RUBA\$TRI	-DAKEIAFDQKSR A--ALQAG-----VEKLANAVGVTLG
P60\$HOMO	-YAKDVKFGADARAL--MLOG-----VDLLADAVAVTMG
P60\$YEAS	-SHKELKFGVEGRAS--LLKG-----VETLAEAVAATLG

. . . * * *

TF55	PRGMDKMFVDSLGDITITNDGATILDKMDLQHP----TGKLLVQIAKGOD
ECOGROE	PKGRNVVLDKSFAGPTITKDGVSVAIEIELEDKFENMGAMVKEVASKAN
MICOBOVI	PKGRNVVLEKKWGAPTITNDGVSIAKEIELEDPYEKIGAEVKEVAKKTD
GROL\$SY3	PKGRNVLLEKQYGAPQIVNDGITVAKIEIELSNPEENAGAKLQEVASKTK
RUBB\$BRA	PKGRNVVLESKYGSPRIVNDGVTVAREVELEDPVENIGAKLVRGAAAKTN
RUBA\$TRI	PRGRNVVLD-EYGNPKVVNDGVTIARAIELANPMENAGAALIREVASKTN
P60\$HOMO	PKGRTVIEEQSWGSPKVTKDGVTVAKSIDLKDKYKNIGAKLVQDVANNTN
P60\$YEAS	PKGRNVLIEQPFPGPKITKDGVTVAKSIVLKD K FENMGAKLLOEVASKTN

* . * . . . * . . . * * . . . * * .

137

TF55	EETADGKTAVILAGELAKKAEDLLYKEIHPTIIVSYKKAEEIALKTIQ
ECOGROEL	DAAGDGT TTTATVLAQAIITEGLKAVAAGMNPMDLKRGI DKA VTA AVEELK
MICOBOVI	DVAGDGT TTTATVLAQALVREGLRNVAAGANPLGLKRGIEKAVEKVTETLL
GROL\$SY3	EIAGDGT TTTATIIAQALVREGLRNVAAGANPVALRRGIEKVTTFVLQGEIE
RUBB\$BRA	DLAGDGT TTSVLAQGFIAEGVKVVAAGANPVLITRGI EKTAKALVAELK
RUBA\$TRI	DSAGDGT TTTACVLAREI I KLGILSVTSGANPVSLKKGIDKTVQGLIEELE
P60\$HOMO	EEAGDGT TTTATVLARSIAKEGF EKISKGANPVEIRRGVMLAVDAVIAELK
P60\$YEAS	EAAGDGT TTSATVLGRAIFTESVKNVAAGCNPMDLRRGSQVAVEKVI EFLS

. . . * * * * . . . *

TF55
ECOGROEL
MICOBOVI
GROL\$SY3
RUBB\$BRA
RUBA\$TRI
P60\$HOMO
P60\$YEAS

DIAQPVSIINDT-----DVLKVALTSLGSKAVAGAREYLA
ALSVPSCSDSKAIAOVGTISANSDETVGKLI AEAMDKV GKEGVITVED--G
KGAKEVETKEQIAATAAISA-GDQSIGDLIAEAMDKV GNEGVI TVEE--S
AVAKPVEGS-AIAQVATVSSG TNPEVGAMIADAMDKVTKDGVITVEE--S
KMSKEVEDSE-LADVAAVSAGNNAEIGSMIAEAMSRVGRKGVV TLEE--G
RKARPVKGSGDIKAVASISAGNDELIGAMIADAIDKVGPDGVLSIES--S
KOSKPVTTPEEIAQVATISANGDKEIGNIISDAMKKVGRKGVITVKD--G
ANKKEITTSEEIAQVATISANGDSHVGLKLLASAMEKVGKEGVITRE--G
* *

TF55
ECOGROEL
MICOBOVI
GROL\$SY3
RUBB\$BRA
RUBA\$TRI
P60\$HOMO
P60\$YEAS

DLVVKAVAQVAELRGDKWYVDLNDVQIVKKHGGSINDTQLVYGIVVDKEV
TGLQDELDDVVEGMOQFDRGYL---SPYFINKPETGAVELES PFILLADKKI
NTFGLQLELLETEGMRFDKGYI---SGYFVTDPERQEAVLEDPYIILLVSSKV
KSLNTELEVVEGMOIDRGYI---SPYFITDSDRQLVEFDNPLILITDKKI
KSAENALYVVEGMOQFDRGYV---SPYFVTDSEKMSVEFDNCKLLLVDKKI
SSFETTVDVEEGMEIDRGYI---SPQFVTNLEKSI VEFENARVLITDOKI
KTLNDELEIIEGKMFDRGYI---SPYFINTSKGQKCEFDAYVLLSEKKI
RTLDELELVTEGMRFDGFI---SPYFITDPKSSKVEFEKPLLLLSEKKI
* *

TF55
ECOGROEL
MICOBOVI
GROL\$SY3
RUBB\$BRA
RUBA\$TRI
P60\$HOMO
P60\$YEAS

--VHPGMP--KRIENAKIALLDASLEVEKPELDAEI--RINDPTQMHKFL-
SNIREMLPVLEAVAKAGKPLLI AEEDVEGEALATAVVNTIRGIVKVAAVK
STVKDLLPLLEKVI GAGKPLLI AEEDVEGEALSTLVVVKIRGTFKSVAVK
SAIAELVPVLEAVARAGRPLLI AEEDIEGEALATLVVVKARGV LNVA AIK
TNARDLVGVLEDAIRGYPILII AEEDIEGEALATLVVVKLRGTLKIAALK
TSIKEIIPLEQTTLRCP LFIVAEDITGEALATLVVVKLRGI INVAAIK
SSIOSIVPALEIANHRKPLVIAEDVDGEALSTLVNRLKVLQVAVK
SSIQDILPALEISNQSRPLLI AEEDVDGEALACILNKLRGQVKVCAVK
* *

TF55
ECOGROEL
MICOBOVI
GROL\$SY3
RUBB\$BRA
RUBA\$TRI
P60\$HOMO
P60\$YEAS

EEE-ENILKEKVDKIAATGANVVICQKGI----DEVAQHLYLAKKGILAVR
APGFGDRRKAMLQDIATLTGGTVISEE-IGMELEKATLEDLQAKRVVIN
APGFGDRRKAMLQDMAILTGGQVISEE-VGLTENADLSLLGKARKVVVT
AP-FGDRRKAVLQDIAILTGGSVISED-IGLSLDTVSLDQLGQAVKATLE
APGFGGERKSQYLDIAILTGATVIREE-VGLSLDKAGKEVLGHAAKVVLT
APSFGERRKAVLQDIAIVTGAEYLAKD-LGLLVENATVDQLGTARKITIH
APGFGDNRKNQLKDMAIATGGAVFGEGLTNLEDVQPHDLGKVGVEVIVT
APGFGDNRKNTIGDIAVLTGGTVFTEE-LDLKPEOCTIENLGS CDSITVT
* *

TF55
ECOGROEL
MICOBOVI
GROL\$SY3
RUBB\$BRA
RUBA\$TRI
P60\$HOMO
P60\$YEAS

RAKSDLEKLARATGGRVINSNIDELTSODLGYAALVEERKVGEDKMFVFE
KDTTTIIDGVGE--EAAIQGRVAQIRQQIEEATSD-YDREKLQERV----
KDETTTVEGAGD--TDAIAGRVAQIRQEIENS DSD-YDREKLQERL----
KDNTILVAGAHKRRSAGVKERIEQLRKEYAASDSD-YDKEKIQERI----
KETSTIVGDGSTQ--DAVQKRVTQIKNLI EQAEQD-YEKEKLNRI----
QTTTTLIADAASK--DEIOARVAQLKKE LSETDSI-YDSEKLAERI----
KDDAMLLKKGDK--AQIEKRIQEIIEOLDVTTSE-YEKEKLNRL----
KEDTVILNGSGPK--BAIQERIEQIKGSDITTTNSYEKEKLQERL----
* *

TF55 GAKNPKSVSILIRGGLERVVDETERALR--DALGTVADVIRDGRAVAGGG
 ECOGROEL -AKLAGGVAV- IKVGAATEVEMKEKKARVEDALHATRAAVEEG-VVAGGG
 MICOBOVI -AKLAGGVAV- IKAGAATEVELKERKHRIEDAVRNAKAAVEEG-IVAGGG
 GROL\$SY3 -AKLAGGVAV- IKVGAATEVELKDRKLRIEDALNATKAAVEEG-IVPGGG
 RUBB\$BRA -AKLGGGVAV- IOVGAOTETELKEKKLRVEDALNATKAAVEEG-IVVGGG
 RUBA\$TRI -AKLGGGVAV- IKVGATTETELEDROLRIEDAKNATFAAIEEG-IVPGGG
 P60\$HOMO -AKLSDGVAV- LKVGGSVSDVEVNEKKDRVTDALNATRAAVEEG-IVLGGG
 P60\$YEAS -AKLGGGVAV- IRVGGASEVEVGEKKDRYDDALNATRAAVEEG-ILPGGG
 * * * * *

TF55 AVEIEIAKRLRKYAPQVGGKEQ-LAIEAYANAIEGLIMILAENAGLDPID
 ECOGROEL VALIRVASKLADLRGK--NEDQNVGIKVALRAMEAPLRQIVLNCGEEP--
 MICOBOVI VTLLOAAPTLDLDELK-L--EGDEATGANIVKVALEAPLQKIAFNSGLEP--
 GROL\$SY3 TTLIRLAGKIESFKAQLSNDEERVAADI IAKALEAPLHQLASNAGVEG--
 RUBB\$BRA CTLRLRLASKVDAIKAALNDDEEKVADIVKRALSYPLKLIAKNAGVNG--
 RUBA\$TRI AAYVHLSTYVPAIKETIEDHDERLGADI IOKALOAPASLIANNAGVEG--
 P60\$HOMO CALLRCIPALDSL--TPANEDQKIGIEI IKRTLKIFAMTIAKNAGVEG--
 P60\$YEAS TALVKASRVLDEV--VVDNFDQKLGVDI IRKAITRPAKQI IENAGEEG--
 * * * * *

TF55 KLMQLRSLHENETNKWYGLNLFNGPDMWKLGVIEPALVKMNAIKAATE
 ECOGROEL SVVANTV-KGGDGNV--GYNAATEEYGNMIDMGILDPTKVTRSALQYAAS
 MICOBOVI GVVAEKV-RNLPAGH--GLNAQTGVYEDLLAAGVADPVKVTRSALQNAAS
 GROL\$SY3 SVIVEKV-KEATGNQ--GYNVITGKIEDLIAAGI IDPAKVRSALQNAAS
 RUBB\$BRA SVVSEKVLSDNDNVKF--GYNAATGKYEDLMAAGI IDPTKVVRCCLEHAAS
 RUBA\$TRI EVVIEKI-KESEWEM--GYNAMTDKYENLIESGVIDPAKVTRCALQNAAS
 P60\$HOMO SLIVEKIMQSSSEV---GYDAMAGDFVNMEKGI IDPTKVVRTALLDAAG
 P60\$YEAS SVIIGKLIDEGDDFAKGYDASKSEYTDMLATGI IDPFKVVRSGLVDAAG
 * * * * *

TF55 AVTLVLRIDDIVAAGKGGSEP-----GKKKEKEEKSSD 552
 ECOGROEL VAGLMITTECMVTDLPKND-ADLGAAGGMGGMGGMGGM---M 548
 MICOBOVI IAGLFLTTEAVVADKPEKEK-ASVPGGGDMGGM-----F 540
 GROL\$SY3 IAGMVLTTTEALVVEKPEPAP-PAMPDMGGMGGMGGMGGM-GMM 551
 RUBB\$BRASI VAKTFLMSDCVVVEIKEPEP-VP----- 534
 RUBA\$TRIGO VSGMVLTTGAIIVVEKPKPKPKVAEPAEGOLS-----V 541
 P60\$HOMO VASLLTTAEVVVTEIPKEEKDPGGMGAMGGMGGM--GMGG--GMF 548
 P60\$YEAST VASLLATTEVAIVDAPEP---PAAAGAGGMPG--GMPGMPGMM 550
 * * * * *

Residuos

Identidad : 39 (6.6%)
 Similaridad : 175 (29.5%)
 Promedio general
 de conservación 214 (36.1%)

Fig 3. Secuencia consenso.

La secuencia consenso, se construyó como una alternativa para corroborar las características generales de la proteína (hidropatía, identidad, distancias evolutivas, relaciones filogenéticas etc). Esta secuencia conserva el comportamiento global de las cpn60 (datos no mostrados), por lo que ha sido utilizada en la predicción de estructura secundaria promedio, así como en la ubicación de periodicidades de residuos. Esta secuencia fue deducida en forma manual atendiendo diferentes criterios en los residuos (como las características fisicoquímicas, tamaño, etc).

AKEIKFXXEA RRXLQAGVDK LADAVKVTLG PKGRNVVLEX KFGSPKITND
GVTVAREIEL EDPFENIGAK LVKEVASKTN DAAGDGTTTA TVLAQAIKE
GLKNVAAGAN PMXLKRGIDK AVXAVIEELK KXAKPVXXSE XIAQVATISA
GNDXXIGXMI ADAMDKVGXE GVITVEEGKS ETDELEVXE GMXFDRGYIS
PYFVTDPEXQ XVEFENAXIL LXDKKISSIX DIVPXLEXVX XXGKPLIIA
EDVEGEALAT LVVNKLGRXL KVAAVKAPGF GDRRKAXLQD IAILTGGXVI
XEELGLXLEX AXLEQLGXAR KVTITKDXTT LVXGXGXKDA VQXR VXQIRX
EIXXTDSYD KEKLQERIAK LSGGVAVIKV GAATETELKE RKLRIEDALN
ATKAAVEEGI VPGGGXALXR LAXKXDAIKX ANXDEKXGAD IVKRALXAPL
XLIAXNAGVE GSVVVEKVXE LXXXGYNAXT GKYENLXAAG IIDPAKVTRS
ALQNAASVAG MVLTTTEAVVW DKPEPEPXPX XXAGGMGGM

Gráfica de puntos de la alineación múltiple

Se hizo tanto en forma manual como con un programa de software y, como se observa en la figura 4, se detectaron 5 regiones de máxima homología, así como otras tantas de hipervariabilidad. Estas zonas se caracterizan en lo general por no presentar ninguna distribución especial, ya que se localizan a lo largo de la proteína, lo mismo en la región amino (21-100), como en el extremo carboxilo (520-525).

La primer región en el extremo amino es la de mayor longitud, se localiza aproximadamente a partir del residuo número 21, y se extiende hasta aproximadamente el aminoácido 100. Esta es una de las dos zonas donde se encuentran regiones con 90% de conservación. Esta región posee un perfil hidropático poco definido, solo al final de ella es claramente hidrofóbica.

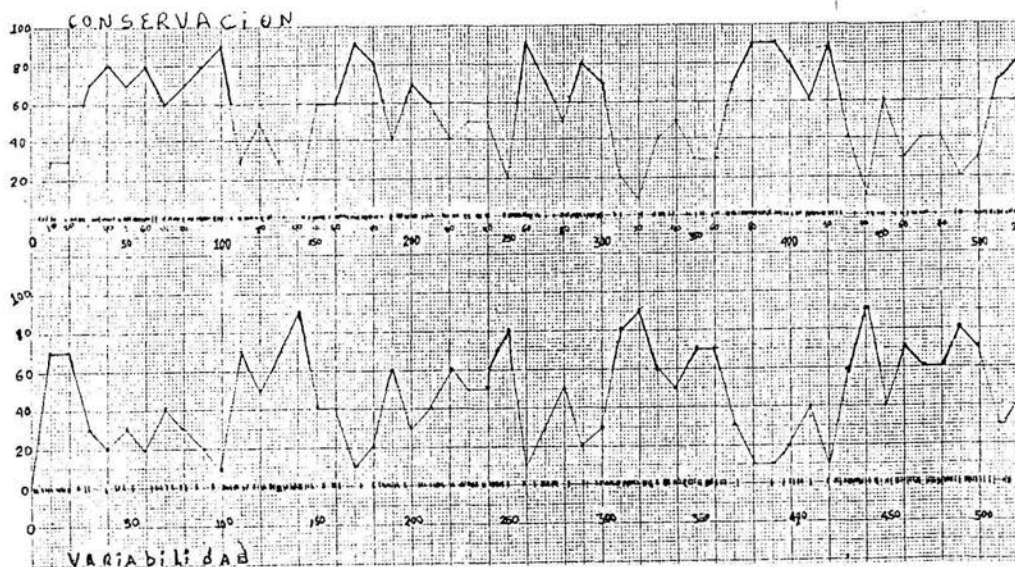
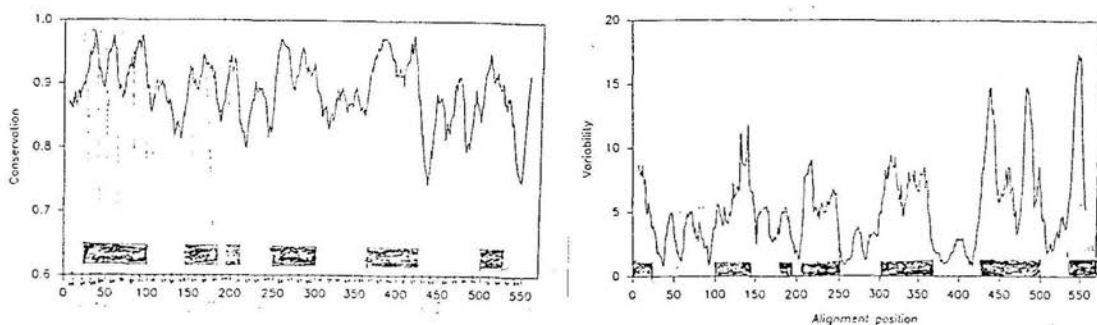
Probablemente presente en su estructura secundaria regiones alternas de alfa hélices y cadenas beta plegadas, en aproximadamente iguales proporciones.

La siguiente región está dividida en dos zonas pequeñas. La primera de ellas empieza en el residuo 146 y termina en el aminoácido 185, posteriormente se encuentra un pequeño intervalo de 9 aminoácidos de escaso valor conservacional, en seguida se encuentra la segunda zona que empieza en el aminoácido número 194 y termina en el residuo 212. Esta región posee un nivel de hidropatía no definido, quizás podría ser una región anfipática. A nivel de estructura secundaria se encuentra la predicción más fuerte para las beta plegadas (posición 200).

La tercer región corresponde a la zona comprendida del residuo número 250 hasta el 300. Esta región prácticamente divide al polipéptido en dos, y su característica es básicamente hidrofóbica.

Es una de las dos zonas con mayor definición en su perfil hidropático consenso. La predicción a nivel de estructura secundaria la revela como una de las regiones con mayor riqueza de alfas hélices, observándose solo una pequeña estructura beta hacia el final de la región. Esta región podría ser la parte central de la proteína alrededor de la cual se estructura el resto de la molécula. La **penúltima región** empieza aproximadamente en el residuo 367, finalizando en el 425. Esta zona es la segunda más grande, contiene 58 residuos, y presenta una región 100% conservada. Esta zona se perfila como la más importante, ya que comparte varias características en común para todas las cpn60 estudiadas. Además de las características ya descritas (alto grado de conservación, y tamaño), posee el más alto grado de conservación a nivel de perfil hidropático, con una tendencia alternada, ya que empieza con una fuerte inclinación hidrofílica, para posteriormente cambiar abruptamente hacia hidrofóbica también de manera muy fuerte. Esta tendencia en forma atenuada continua hasta el final de la región. Además, la predicción de estructura secundaria muestra una zona con elevada proporción de alfas hélices, y también aquí se observa una de las predicciones mas consistentes para las beta plegadas hacia el final de la región. La **quinta región** abarca aproximadamente del residuo 500 al 525. Es la zona más cercana al carboxilo terminal, y en su promedio hidropático muestra una fuerte tendencia hacia la hidrofobicidad. Su predicción de estructura secundaria revela una mayor inclinación hacia la vuelta, con una pequeña alfa hélice. Esta zona ha sido implicada en la interacción con GroES. Sin embargo, las evidencias son aun muy escasas, ya que Langer *et al* (1992), deletaron hasta 50 residuos del extremo carboxilo de GroEL sin que ello afectara la actividad de unión de GroES/GroEL.

Fig 4. Gráfica de puntos, Muestra las diferentes regiones de conservación (izquierda), y de variabilidad (derecha). Además de estas figuras, las zonas fueron también determinadas independientemente en forma manual, como se observa en las figuras inferiores, en donde al pie de ellas se pueden advertir líneas densas que representan a cada posición de los residuos conservados, tanto en la gráfica de variabilidad como la de conservación. En general, se aprecia que las gráficas resultaron similares. La zona 1 inicia en el residuo 21 y termina en el 100, la zona 2, esta dividida en dos subregiones, la primera inicia en el residuo 146 y termina en el 185, la segunda empieza en el aminoácido 194 y finaliza en el 212; la tercer región empieza en el residuo 250 y termina en el 300; la cuarta zona principia en el residuo 367 y finaliza en el 425; la última zona comienza en el residuo 500 y termina 525.



Agrupamiento de residuos

En el comando PRESIDUE del Pcgene, se realizó una extensa búsqueda de agrupamientos de los diferentes grupos de aminoácidos (hidrófobos, polares, aromáticos, residuos individuales, etc), utilizando como modelo a la secuencia consenso, de estos solo resultaron positivos, la prolina como residuo individual (fig 5a).Y los grupos aromáticos (fig 5b).

Fig 5a. Distribución de prolinas altamente conservadas.

En esta gráfica se distinguen 14 prolinas conservadas, de las cuales cinco pudieran ser importantes en la estructura secundaria como rompedoras de alfas hélices. Estas prolinas estarían en las posiciones 35,49,250,280 y 494. En el eje horizontal se muestra la posición de los residuos en la proteína, y en el eje vertical la frecuencia de los residuos a cuantificar (ver fig 7).

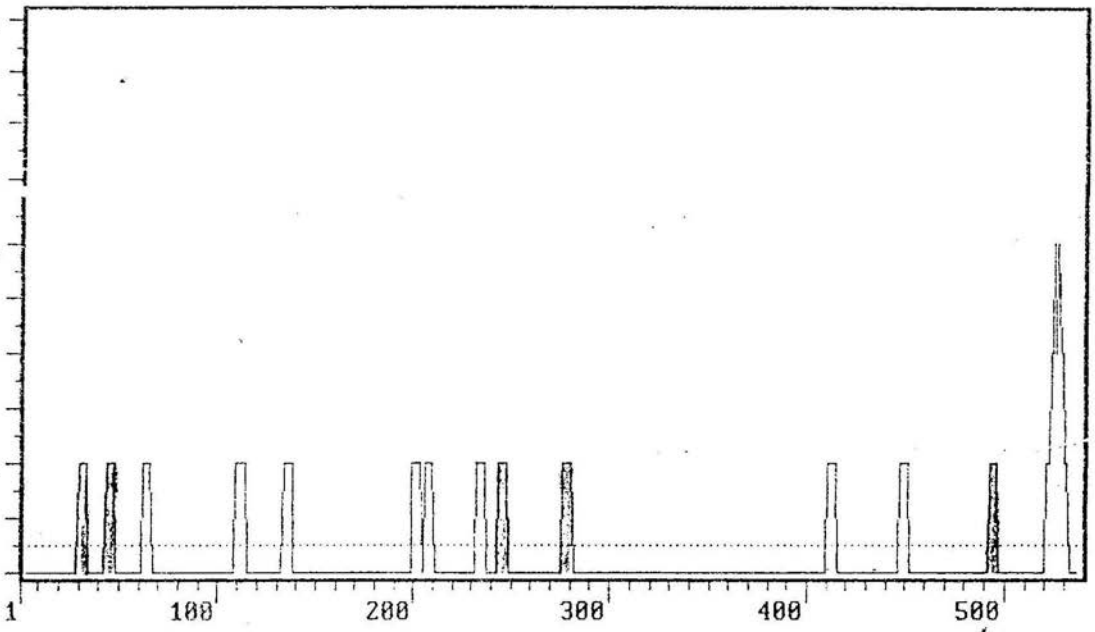


Fig 5b. Localización de residuos aromáticos. Como se observa en esta figura, se localizaron 6 residuos aromáticos, cuyas posiciones serían las siguientes; 8,42,202,207,284 y 367. El eje horizontal muestra la posición de los residuos a cuantificar en el total de la proteína, el eje vertical indica la frecuencia de aparición de los residuos.

La importancia de estos agrupamientos no es clara, solo es posible que, la última posición pueda tener alguna trascendencia en las reacciones cooperativas de los cambios conformacionales, en el mecanismo de auxilio del plegamiento de las proteínas mediado por chaperoninas propuesto por Jackson. *et al* (1993).

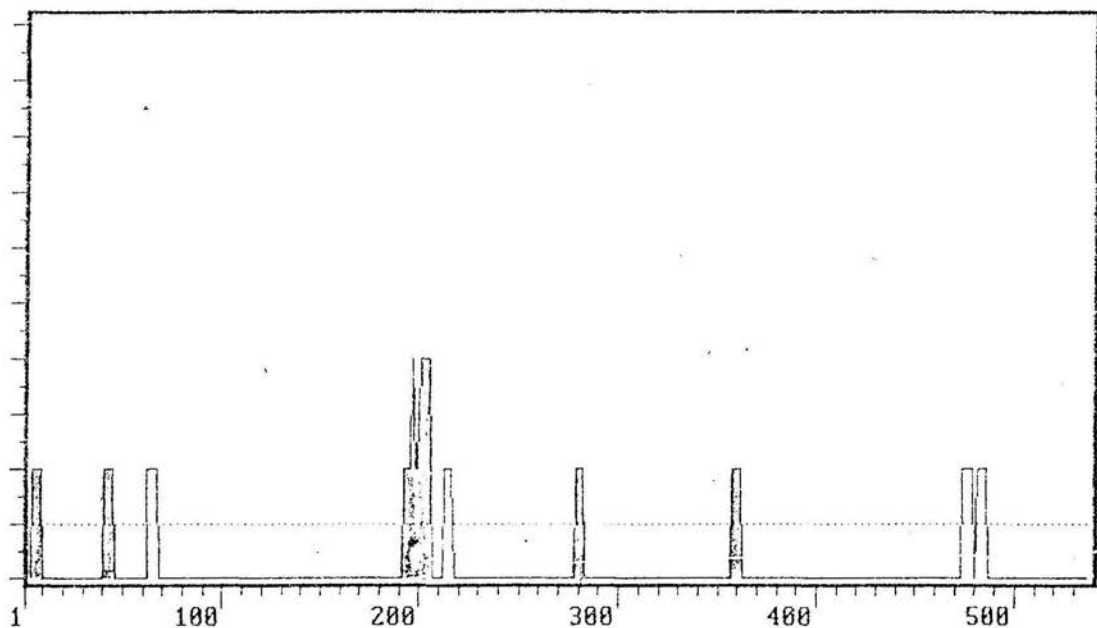


Fig 6. Perfil hidropático promedio.

En el eje horizontal se cuantifica el tamaño de la proteína, en el eje vertical se puede notar dividido en dos secciones, en la parte superior se cuantifican regiones hidrofóbicas, mientras que en la sección inferior se grafican los caracteres hidrofílicos.

Como se puede observar es notable las alternancia de los caracteres a partir de aproximadamente el residuo número 270 (especialmente en la zona que corresponde a la cuarta región de alta conservación residuo 367-425), hasta el final de la secuencia. Se nota además que la proteína tiende más a la hidrofobicidad desde el inicio hasta aproximadamente el residuo 270.

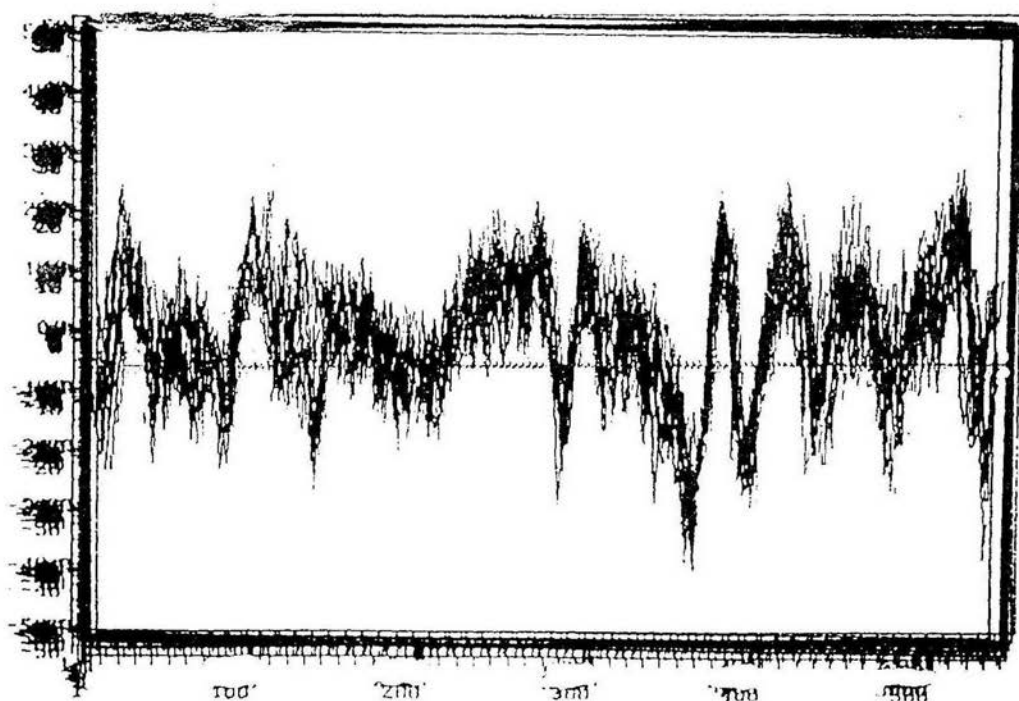
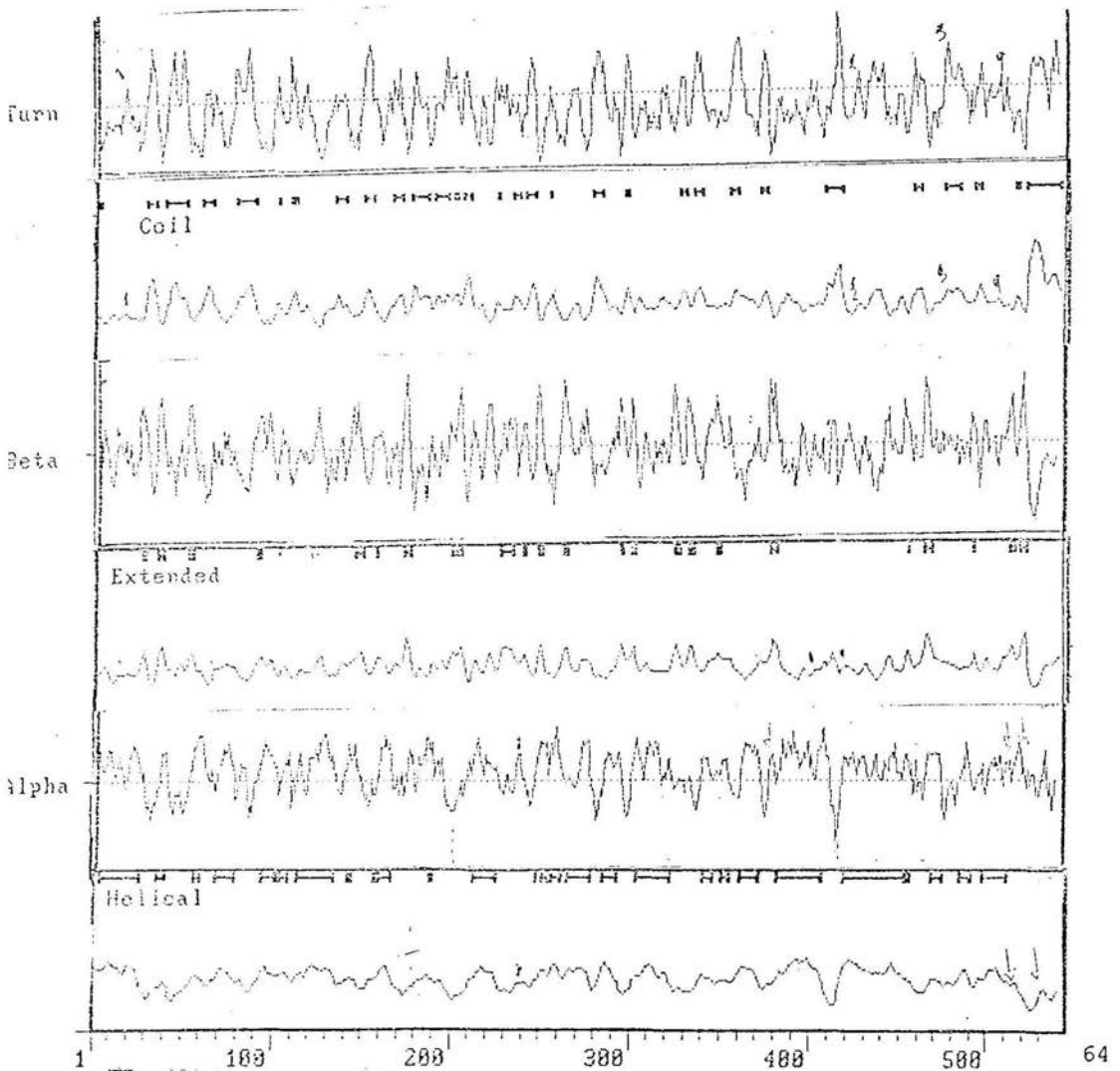


Fig 7. Predicción de estructura secundaria consenso.

En esta gráfica se observa que la proporción de alfas hélices es cercana a la mitad de la proteína (50.6%), siendo las beta extendidas segundo lugar con (33.5%), ocupando las vueltas el tercer sitio (15.7%). Los caracteres alfa/beta se encuentran en forma alterna, por lo que se le puede clasificar a este péptido como alfa-beta. Es importante advertir que estas predicciones son parecidas a los resultados de los estudios por dicroísmo circular realizados por Lissin, *et al* (1992), en el cual reporta hasta un 60% de alfas hélices en la estructura secundaria de cpn60. Esta gráfica fue construida en base a 3 diferentes métodos.



Arboles filogenéticos.

Fig 8a. Arbol 1. Esta figura fue creada en base a la segunda etapa de los alineamientos múltiples (comando CLUSTAL del Pcgene), en el se utilizan los cálculos basados en el método UPGMA, aquí se muestran las relaciones filogenéticas de diferentes especies en la proteína cpn60 (no incluye a las arqueobacterias), estas se cuantifican en base a las distancias evolutivas, se presentan en forma de nódulos, ellos dividen las ramas del dendograma, sus valores están en forma ascendente, (de la relación más cercana a la más alejada). En esta figura se nota claramente el aglutinamiento de los grupos en forma natural. Se observan 13 nódulos que definen a las relaciones entre los grupos que representan a las mitocondrias (mamíferos,levadura), el del grupo de las Eubacterias. Asimismo se nota claramente el estrecho vínculo entre los organismos que representan a los cloroplastos en el grupo de plantas (subunidades alfa y beta) (ver nombres de los organismos en material y métodos).

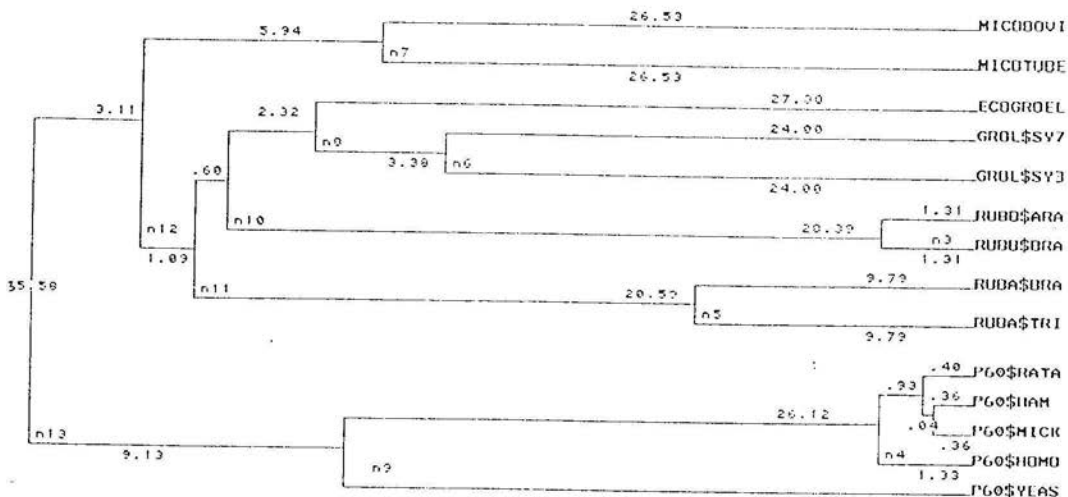


Fig 8b. Arbol 2. Ilustra las relaciones filogenéticas de la proteína cpn60 de diferentes grupos filogenéticos (incluyendo arquea). En esta figura se nota claramente la relación que conserva la proteína cpn60 en los diferentes organismos que representan a los tres linajes primarios (Arqueobacteria, Eubacteria y Eucarionte) propuestos por Woese *et al* (1990). Asimismo se observa que el agrupamiento de este conjunto es en forma natural (los cálculos están basados en el método UPGMA de la segunda etapa de el comando CLUSTAL). TF55, arqueobacteria; ECOG, eubacteria; MICO, eubacteria; GROL, eubacteria; RUBB,RUBA, plantas, eucarionte P60\$, mamíferos-levadura, eucarionte.

Arbol 2; muestra las relaciones filogenéticas de diferentes especies en la proteína cpn60 (incluye archeobacteria). Los cálculos están basados en el método UPGMA

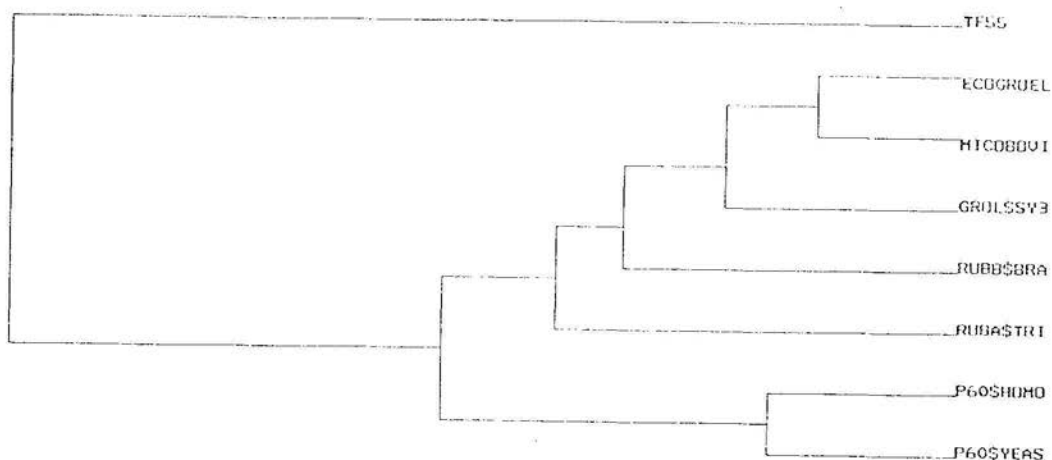
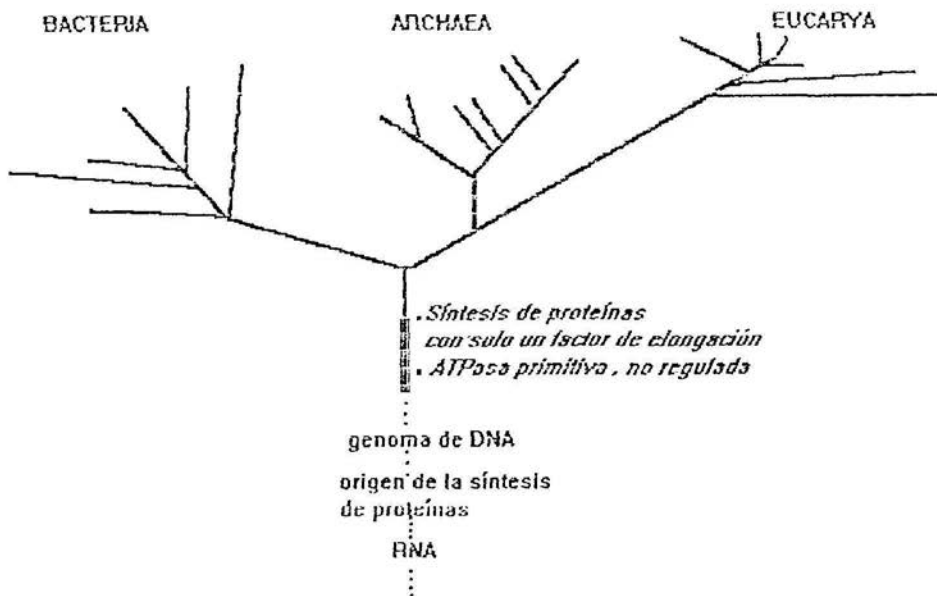


Fig 9 Arbol universal propuesto por Woese *et al* (1990), modificado por Lazcano *et al* (1992).

En este árbol se presentan los 3 dominios filogenéticos universales propuestos por Woese *et al* (1990), basados en el análisis de la comparación de las secuencias de RNA ribosomal. Esta figura fue modificada por Lazcano *et al* (1992), donde el árbol es enraizado utilizando características homólogas comunes de sus descendientes (genes paralogos). Además se ilustran las posibles características para el último ancestro común, así como la posición relativa de los eventos más importantes en la evolución celular, ambas son propuestas hechas por Lazcano *et al* (1992).



7 DISCUSION DE RESULTADOS

Características del progenote

Métodos basados en isótopos radiactivos han dado resultados para la edad de tierra de al menos 4.6×10^9 años, asimismo se han detectado rocas sedimentarias de entre 3.2 a 3.5×10^9 años de antigüedad, en ellas se han encontrado microfósiles parecidos a bacterias, ellos muestran alto grado de complejidad, por lo que es posible que la vida apareciera antes de esta etapa en formas más sencillas. Y es factible que la vida emergiera rápidamente en el planeta. Las evidencias teóricas y experimentales de que la vida no fue un proceso lento es demostrada por la síntesis química orgánica abiótica, en condiciones naturales y experimentales esta se lleva a cabo en períodos cortos.

Las propiedades de un organismo ancestral pueden ser inferidas de la distribución de características homólogas entre sus descendientes, es razonable asumir que, los genes encontrados en organismos de los 3 linajes principales estaban también presentes en el último ancestro común. Lazcano *et al* (1992) describen parcialmente a el último ancestro común, cuyas características serían las siguientes:

1) Aparato traductor con RNAPol oligomérica, ribosomas complejos con diferentes proteínas ribosomales y factor de elongación modulando la síntesis de proteínas. 2) Membrana asociada a actividad ATPasa, involucrando transporte activo (incluyendo mecanismos regulatorios).

- 3) Complejo juego de enzimas, incluyendo aquellas involucradas en la glucólisis, oxidación del piruvato y otras rutas anaeróbicas heterotróficas.
- 4) Enzimas involucradas en la biosíntesis de histidina, sugiriendo la existencia de proteínas catalíticas en las cuales este aminoácido estuvo presente.

No existe actualmente alguna especie con todas las características del estado ancestral. Es lógico suponer que el ancestro común presentaría características metabólicas más simples que las contemporáneas procariontes. Woese *et al* (1987) han definido como progenote a un sistema primitivo de vida, en este, la relación entre el fenotipo y el genotipo es imprecisa, es decir que el acoplamiento entre genoma, transcripción y traducción pudo haber sucedido en solo un evento a nivel muy simple, es claro la falta de una estructura genómica funcional, desde este punto de vista es obvio que el último ancestro común (antes descrito) tenía relativa complejidad, además las similitudes mostradas en las proteínas encontradas en los tres linajes, sugieren que el sistema genético poseía una alta fidelidad. Ello sugiere que el progenote descrito por Woese podría en realidad describir a algún etapa primitiva de la vida dependiente de RNA (Lazcano *et al.* 1992), dada la capacidad catalíticas y de templado del RNA, ella podría desarrollar ambas funciones genotípicas y fenotípicas.

Por lo anterior es prematuro decidir si el último ancestro común descrito en la filogenia puede ser reconocido como un progenote o un verdadero genote.

El caso de las chaperoninas

Como ya se discutió arriba, de acuerdo con los índices (40%-50%) de identidad de todas las cadenas de cpn60 estudiadas (fig 1f), se puede apreciar que son homólogas. Esta conclusión se ve corroborada con los datos de las distancias evolutivas (20%-30%) (fig 1g). Ello significa evolutivamente que el gen que codificaba para esta proteína se encontraba ya presente en el último ancestro común, es decir, que su origen ocurrió antes de su diversificación. Se puede apreciar que su tasa de cambio es baja (20%-30%), lo que sugeriría que su reloj molecular es lento, ya que como se sabe las proteínas tienden a cambiar inicialmente en su estructura primaria, protegiendo los sitios activos de la estructura secundaria y terciaria, por lo que los cambios detectados están en la secuencia de aminoácidos, ya que este péptido está altamente conservado funcionalmente (Lubben *et al.*1989; Lubben *et al.*1990; Gatenby *et al.*1985)

La presencia de chaperoninas en eubacterias y organelos celulares ha sido interpretada como una evidencia del origen endosimbiótico de las mitocondrias y los cloroplastos (Gupta *et al.*1989). En este sentido la comparación de secuencias de citocromo c mitocondrial y RNA ribosomal 16s indican que, las mitocondrias probablemente desciendan de miembros de las bacterias púrpura (Gray y Doolittle,1982; Woese *et al.*1980; Yang *et al.*1985). Asimismo, los datos de las comparaciones de las secuencias de RNA ribosomal también muestran que los cloroplastos probablemente desciendan de las cianobacterias (Gray y Doolittle,1982; Wallace,1982; Gray *et al.*1984).

Mucha información sobre el origen de esos organelos ha sido derivada de esta forma. Sin embargo, son necesarias otras evidencias que refuercen la hipótesis de la endosimbiosis, y la deducción del origen filogenético de los endosimbiontes, por lo que en este trabajo se presentan evidencias adicionales, como la homología de las proteínas cpn60 en bacterias, organelos de eucariontes y probablemente arquabacteria.

Como ya se ha visto en la fig 1h, los promedios de similitud de las subunidades alfa y beta de cloroplastos con las cpn60 de cianobacterias se ubican entre 69.5% y 73.8%. Estas cifras son superiores a los promedios encontrados de estos polipéptidos con los de procariontes no fotosintéticos (66.8%). Este hecho, refuerza la idea de el origen endosimbótico de los cloroplastos, y además vincula a las dos subunidades alfa y beta. Y consecuentemente también convalida la idea del origen común entre las dos subunidades.

En relación a el origen endosimbótico de las mitocondrias, como ya se ha mencionado, se cree que es a partir de bacterias púrpuras (como es el grupo de *Escherichia coli*), y como lo muestra el resultado de nuestras alineaciones fig 1h, donde Ecogroel (*Escherichia coli* y P60\$yeast (*Sacharomyces cerevisiae*) exhiben los más altos índices de similitud (69.5%), en relación a cualquier otro grupo. Los resultados presentes aquí refuerzan la hipótesis del origen endosimbótico de las mitocondrias.

Estructura y función

Las chaperoninas cpn60 (GroEL) actúan cooperativamente con otro tipo de chaperoninas como las cpn10 (GroES). El mecanismo por el cual ejercen su función de molécula chaperona es parcialmente conocido, y se basa principalmente en el análisis de las propiedades estructurales y funcionales de la interacción GroEL/GroES (holochaperonina).

Langer *et al* (1992), observaron estas interacciones al microscopio electrónico y concluyeron que GroES se une en el extremo de uno de los cilindros GroEL, y que alcanza su máxima actividad cuando la relación estequiométrica es 1:1, esta interacción también activa cambios conformacionales del cilindro opuesto, de esta manera aparentemente impide la asociación de un segundo oligómero GroES.

Se ha sugerido que la unión del sustrato proteico a plegar ocurre dentro de la cavidad de GroEL (Braig *et al.* 1993). Esta unión en el interior de la cavidad provee la función esencial de chaperon, ya que restringe a otras moléculas de interaccionar con el polipéptido, evitando así la formación de agregados, además se ha propuesto que esta unión causa rompimiento de enlaces no covalentes en los intermediarios compactos.

En principio en este espacio se pudieran acomodar intermediarios del plegamiento para una proteína monomérica de hasta 90 Kda. Los monómeros mayores pudieran ser unidos en dominios o quizás haya otros factores (Langer *et al.*1992). Se sabe que en GroEL de *Escherichia coli* --

se unen una o dos moléculas de proteína no plegada dentro de su cavidad (Braig *et al.* 1993; Langer *et al.* 1992), probablemente en la conformación de un intermediario compacto (Martin *et al.* 1991).

Se ha demostrado, al menos para *Escherichia coli* (Horwich *et al.* 1993) que las chaperoninas no son específicas para los polipéptidos cuyo replegamiento asisten *in vitro*, ni tampoco que ellas incrementen la tasa de plegamiento, solo auxilian que este sea correcto. La cuestión entonces es:

¿Como es posible para las chaperoninas asistir el plegamiento de una amplia variedad de proteínas frecuentemente muy diferentes, y por que son requeridos dos componentes?

La isomerización de las proteínas de un estado no plegado (U) a las formas activas o nativas (N), usualmente involucran la formación transitoria de intermediarios del plegamiento (I).

La forma U es altamente inestable, por lo que rápidamente transita a una forma termodinámicamente mas estable I. Es probable que este estado tenga un gran numero de residuos hidrofóbicos expuestos en su superficie, y consecuentemente las proteínas en esta forma son menos solubles que en su estado nativo N, y por lo tanto son muy propensas a formar agregados. Así, se ha propuesto que las chaperoninas reconozcan a las proteínas a plegar.

Además es evidente que las proteínas en su estado nativo N, no parecen interactuar con las chaperoninas, lo que sugiere que los motivos estructurales que imparten reconocimiento a cpn60, son inaccesibles o no están presentes. En consecuencia esos motivos podrían ser accesibles solo en la forma de intermediario I.

La exposición de los residuos hidrofóbicos es una característica importante de los intermediarios del plegamiento, y quizás ello permita a las chaperoninas reconocer al sustrato a plegar. Adicionalmente se ha sugerido que las proteínas inapropiadamente plegadas son reconocidas por las cadenas polares expuestas excesivamente desplegadas, en lugar de regiones hidrofóbicas (Hubbard y Sander,1991).

El papel de GroES en la modulación de la función de GroEL es solo parcialmente conocido.

GroES esta presente *in vivo* y es esencial para la viabilidad celular, y es probable que este involucrado con GroEL en el plegamiento de las proteínas en la célula. Cual es el papel exacto de GroES?.

Se han sugerido para GroES tres funciones básicas en la viabilidad celular (Martin *et al.*1993), las cuales son las siguientes:

- 1) Estabiliza uno de los dos anillos de GroEL para darle alta afinidad para ADP. Este es el estado donde se acepta a los polipéptidos parcialmente plegados, la inhibición de la hidrólisis de ATP cuando la proteína no es plegada, reduce el desperdicio de ATP.

2) El reacomodo de GroEL para el intercambio de ADP con ATP incrementa la cooperatividad de subsecuentes hidrólisis de ATP.

El descubrimiento de que GroES, así como GroEL unan ATP puede sugerir que GroES lo done directamente a GroEL como parte del incremento de su cooperatividad con GroEL.

3) GroES vuelve a unirse a GroEL cuando este tiene al polipéptido unido, esta nueva asociación evita al polipéptido ser liberado al medio antes de que sea plegado totalmente.

Con estas y otras evidencias, Martin *et al* (1993), propusieron un mecanismo que implica la interacción repetida de ambas chaperoninas, en un fenómeno dependiente de Mg/ATP. Esta interacción es cíclica hasta que el polipéptido alcanza su forma nativa. así el modelo sería como sigue:

a) GroES se asocia asimétricamente a GroEL en uno de los anillos confiriéndole alta afinidad para ADP. Así el anillo opuesto presenta baja afinidad para el nucleótido, la entrada de la proteína a plegar dispara la disociación de ADP en ambos anillos y activa la liberación de GroES.

b) En ausencia de nucleótidos el substrato se une fuertemente a GroEL y la afinidad para GroES es reducida, ausente el ATP el complejo es estable.

c) Con la entrada de ATP, se debilita la interacción entre GroEL y el polipéptido, GroES se reutiliza en el anillo alterno de GroEL, produciendose un complejo ternario,

d) Simultáneamente o en etapas consecutivamente rápidas, la hidrólisis cooperativa de ATP en ambos anillos causa la liberación de la proteína, permitiendo su plegamiento en la cavidad de GroEL.

e) Si el plegamiento es incompleto después de la hidrólisis de ATP, el polipéptido se vuelve a unir a GroEL, conteniendo ADP y GroES, continuando el ciclo de plegamiento, cuando este finaliza, los polipéptidos que han sido plegados obstruyen sus sitios de afinidad para GroEL saliendo de la cavidad.

Con base en estas evidencias, la chaperona puede poseer al menos dos regiones consenso. Una de ellas sería la región que le confiere reconocimiento a los sustratos a auxiliar, la otra sería la zona en la cual se llevan a cabo las reacciones cooperativas de hidrólisis de ATP. Adicionalmente se proponen otras regiones, estas también serían importantes, una de ellas sería la región en la cual interactúan cada uno de los cilindros de la estructura cuaternaria, otra zona sería aquella en la cual interaccionan las subunidades individuales. Además podría existir otra región, cuya función podría ser como sitio de unión a la cochaperonina (GroES).

En este trabajo se describen 5 zonas que hemos catalogado de alta conservación para las chaperoninas, las cuales podrían cumplir algunas de esas funciones.

Se ha sugerido que la chaperonina presenta dos conformaciones, una en la cual se une fuertemente a la proteína no nativa, y débilmente a Mg-ATP, otra en la cual se une firmemente a Mg-ATP, pero débilmente al sustrato.

En base a ello, Saibil y Wood (1993) sugieren para las chaperoninas un modelo anfipático. En este sentido, Jackson *et al* (1993), han propuesto que el mecanismo de plegamiento transita por estas dos conformaciones, es así que el sustrato circula transitoriamente en la cavidad de GroEL, el cual alterna alta y baja afinidad por dicho sustrato. Ellos proponen que la función de los nucleótidos es proporcionar la energía para mantener esos cambios, hasta que la proteína alcanza su estado nativo, y es liberada.

Por otra parte Martin *et al* (1993) utilizando azido-ATP marcado, revelaron un dominio de unión a los nucleótidos de 40K (residuos 153-531), usando la misma técnica detectó una tirosina altamente conservada, sugiriendo que este residuo tiene relación a la unión de los anillos de purina de los nucleótidos. En base al análisis de residuos altamente conservados realizado aquí, proponemos que este aminoácido se encuentra en la posición número 367, la que corresponde a la cuarta zona de alta conservación, en la cual la tirosina solo es sustituida por otro residuo aromático (triptofano) en *Micobacterium tuberculosis* (micotube). Ello sugirió a la cuarta zona de conservación, como el sitio donde probablemente se llevarían a cabo las reacciones cooperativas.

8 CONCLUSIONES

Relaciones evolutivas.

La presencia de proteínas relacionadas a las chaperoninas en los tres linajes primarios (eucaria, arquea y eubacteria), indica de forma convincente que una forma ancestral de esta proteína "chaperonina ancestral", estuvo también presente en el último ancestro universal.

La aparición muy temprana de una proteína relacionada a las chaperoninas, probablemente indica que esta desempeñó una función fundamental, vital para la forma de vida temprana. La función de este grupo de proteínas como "moléculas chaperonas", asistiendo el plegamiento de cadenas polipéptidicas, así como su ensamble dentro de estructuras oligoméricas debió de haber sido fundamental para esas condiciones de vida.

El último ancestro común probablemente muy primitivo, presentaba solo funciones muy esenciales y en forma rudimentaria, en este contexto la presencia de una proteína relacionada a las chaperoninas, pudiera haber desempeñado un papel plataforma, coadyuvando en forma general en estos mecanismo metabólicos, así como en su evolución hacia formas estructurales y funcionales más complejas.

Por otra parte la función de las chaperoninas como proteínas de choque térmico, debieron quizás haber desempeñado un importante papel en la protección de las formas de vida primitivas (previniendo daños ocasionados por las altas temperaturas prevalecientes).

Los resultados sugieren un origen evolutivo común. Sin embargo, estos mismos datos también insinúan que, en contraste a el origen endosimbiótico de las chaperoninas de cloroplasto y mitocondria (vía eubacterias), la cpn60 de arqueabacteria pudieron haber descendido directamente del último ancestro universal común, y consecuentemente divergido de manera diferente.

La cercana relación observada en la similitud de las chaperoninas de cloroplasto y mitocondria, con las de eubacterias aporta una evidencia adicional sobre el origen, y la evolución de estos organelos, especialmente como un reflejo del origen evolutivo de ellos a través de eubacterias, hecho que se estima ocurrió aproximadamente al menos hace 1.9×10^9 millones de años (Gray y Doolittle, 1982).

Estructura y función.

En base a los resultados anteriores, podemos concluir que de las 15 cpn60 analizadas, el 51.6% de los aminoácidos se conservan en forma general, y en particular los residuos importantes en la predicción de la estructura y función (tirosina y prolina), además existen regiones donde se puede apreciar hasta el 90% de conservación, una de las cuales coincide con el hecho de los perfiles hidropáticos muy conservados. La estructura secundaria pertenece a la categoría alfa-beta según las predicciones de las características y propiedades estructurales de esta proteína.

Por otra parte, se cree que las chaperoninas deben poseer al menos 5 zonas estructural y funcionalmente conservadas. Desde este punto de vista el descubrimiento de las zonas conservadas en la estructura primaria de las chaperoninas podría ser relevante, dada la importancia del papel de estas proteínas.

En este trabajo se ha propuesto que la cuarta de estas zonas sea el sitio donde se lleva a cabo la conversión de conformaciones. Esto se basa en las características muy especiales que presenta esta región. A las restantes zonas no es posible asignar ninguna función, ni estructural ni funcional, por lo que queda aquí la simple descripción física de ellas.

9 REFERENCIAS.

- 1.- Ang, D., Liberek, K., Skowrya, D., Zylicz, M., Georgopoulos C: Biological role and regulation of the universally conserved heat shock proteins. *J. Biol. Chem.* 266: 24233-24236 (1991).
- 2.- Anathan, J., Goldberg, A.L., Voellmy, R: Abnormal proteins serve as Eukaryotic stress signals and trigger the activation of heat shock genes. *Science.* 232, 522-524 (1986).
- 3.- Barraclough, R., Ellis, R.J: Protein synthesis in chloroplasts. IX. Assembly of newly synthesized large subunits into ribulose biphosphate carbixylase in insolated pea choroplasts. *Biochim. Biophys. Acta.* 608:19-31 (1980).
- 4.- Beckman, R.P., Mizzen, L.A., Welch, W.J: Interaction of hsp70 with newly synthesized proteins: Implications for protein folding and assembly. *Science.* 248, 850-854 (1990).
- 5.- Bloom, M.V., Milos, P., Roy, H: Light-dependent assembly of ribulose 1,5 bisphosphate carboxylase. *Proc. Natl. Acad. Sci. USA.* 80: 1023-1017 (1983).
- 6.- Bochkareva, E.S., Lissin, N.M., Girshovich, A.S: Transient association of newly synthesized unfolded protein with the heat Shock GroEL protein. *Nature.* 336: 254-257 (1988).
- 7.- Braig, K., Simon, M., Furuya, F., Hainfeld, J.F., Horwich, A.L: A polypeptide bound by the chaperonin groEL is localized within a central cavity. *Proc.Natl.Acad.Sci. USA.* 90, 3978-3982 (1993).
- 8.- Brenner, S: That lonesome grail. *Nature.* 358, 27-28 (1992).

- 9.- Chandrasekhar, G.N., Tilly, K., Woolford, C., Hendrix, R., Georgopoulos C: Purification and properties of the groES morphogenetic protein of *Escherichia coli*. *J. Biol. Chem.* 261: 12414-12419 (1986).
- 10.- Cheng, M.Y., Hartl, F.U., Martin, J., Pollock, R.A., Kalousek, F., Neupert, W., Halberg, E.M., Halberg, R.L., Horwich, A.L: Mitochondrial heat shock protein hsp60 is essential for assembly of protein imported into yeast mitochondria. *Nature.* 337: 620-625 (1989).
- 11.- Chiang, H.L., Terlecky, S.R., Plant, C.A., Dice, J.F: A role for a 70-kilodalton heat shock protein in Lysosomal degradation of intracellular proteins. *Science.* 246, 382-385 (1989).
- 12.- Chirico, W.J., Waters, M.G., Blobel, G: 70K heat shock related proteins stimulate protein translocation into microsomes. *Nature.* 332, 805-809 (1988).
- 13.- Chitnis, P.R., Nelson, N: Molecular cloning of the genes encoding two proteins of the Cyanobacterium *Synechocystis* sp. PCC 6803. *J.Biol.Chem.* 266, 58-65 (1991).
- 14.- Dayhoff, M.O., Atlas of protein sequence and structure 5. *National Biomedical Res. Foundation.* Washington. D.C (1978).
- 15.- DeLuca-Flaherty, C., McKay, D.B., Perham, P., Berth, L.H: Uncoating protein (hsc70) binds a conformationally labile domain of clathrin light chain LCa to stimulate ATP hydrolysis. *Cell.* 62, 875-877 (1990).
- 16.- Dice, J.F: Peptide sequences that target cytosolic proteins for lysosomal proteolysis. *Trends. Biochem. sci.* 13, 305-309 (1990).
- 17.- Doolittle, R.F. Of Urfs And Orfs: A primer on how to analyze derived amino acid sequences. *Library of congresso. U.S.A.* Pag cons. 3,10,53,55 y 56 (1986).

- 18.- Ellis, R.J., Hemmingsen, S.M: Molecular chaperones: Proteins essential for the biogenesis of some macromolecular structures. *Trends. Biochem. Sci.* **14**, 339-342 (1989).
- 19.- Ellis, R.J, Van der Vies S.M: Molecular chaperones. *Annu. Rev. Biochem.* **60**: 321-347 (1991).
- 20.- Fayet, O., Louarn, J-M., Georgopoulos C: Supression of the *E. coli dnaA46* mutation by amplification of the *groES* and *groEL* genes. *Mol. Gen. Genet.* **202**:435-445 (1986).
- 21.- Fayet, O., Ziegelhoffer, T., Georgopoulos C: The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bact.***171**: 1379-1385 (1989).
- 22.- Frydman, J., Minmergern, E., Bromage, H.E., Wall, S.J., Tempt, P., Hartl, F.V: Function in protein folding of Tric, a cytosolic ring complex containing TCP-1 and structurally related subunits: *EMBO.J.* **11**, 4767-4778 (1992).
- 23.- Gao, Y., Thomas, O., Chow, R.L., Lee, G.H., Cowan, NJ: A cytoplasmic chaperonin that catalize beta-actina folding. *Cell.* **69**, 1043-1050 (1992).
- 24.- Garnier, J.D., Osgusthorpe, J., Robson, B: Analysis of the accuracy and implications of simple methods for precting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97- 120 (1978).
- 25.- Gascuel, O.Y Goldmard, J.L: A simple method for prediction the secondary structure of globular proteins. *CABIOS.* **4**, 357-365 (1988).

- 26.- Gatenby, A.A., Lubben, T.H., Ahlquist, P., Keesgstra, K: Imported large subunits of ribulose biphosphate carboxylase/Oxygenase, but not imported beta-ATP synthase subunits, are assembled into holoenzyme in insolated chloplast. *EMBO. J.* 7:1307-1314 (1988).
- 27.- Gatenby, A.A., Van der vies, S.M., Bradley, D: Assembly in *E. coli* of a functional multi-subunit ribulose biphosphate carboxylase from a blue-green alga. *Nature.* 314: 617-620 (1985).
- 28.- Gatenby A.A, Ellis RJ: Chaperon function: the assembly of ribulose biphosphate carboxylase- oxygenase. *Annu. Rev. Cell. Biol.* 6:125-149 (1990).
- 29.- Georgopoulos, C.P., Hendrix, R.W., Gasjens, S.R., Kaiser, A.D: Host participation in bacteriophage lambda head assembly. *J. Mol. Biol.* 76, 45-60 (1973).
- 30.- Gething, M.L., Sambrook, J: Protein folding in the cell. *Nature.* 355, 33-45 (1992).
- 31.- Goloubinoff, P., Gatenby, A.A., Lorimer, G.H: GroEL heat shock proteins promote assembly of foreign prokaryotic ribulosa biphosphate carboxylase oligomers in *Escherichia coli*. *Nature.* 337: 44-47 (1989).
- 32.- Goloubinoff, P., Christeller, J.T., Gatenby, A.A., Lorimer, G.H:Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg- ATP. *Nature.* 342, 884-889 (1989).
- 33.- Govezensky, D., Greener, T., Segal, G., Zamir, A: Involvement of GroEL in *nif* gene regulation and nitrogenase assembly. *J. Bact.* 173: 6339-6346 (1991).
- 34.- Gray, M.W., Dolittle, R.F: Has the endosymbiont hypothesis been proven?. *Microbiol. Rev.* 46, 1-42 (1982).

- 35.- Gray, M.W., Sankoff, D., Cedergren, R.J: On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucleic. Acids. Res.* 12, 5837-5852 (1984).
- 36.- Grimm, R., Speth, V., Gatenby, A.A., Schafer, E: GroEL related molecular chaperones are present in the cytosol of oat cells. *FEBS. Lett.* 286: 155-158 (1991).
- 37.- Gupta, R.S: Sequence and structural homology between a mouse T-complex protein TCP1 and the chaperonin family of bacterial (GroEL,60-65 Kda heat shock antigen) and eukaryotic proteins. *Biochem. Int.* 20:833-841 (1990).
- 38.- Gupta, R.S., Picketts, D.J., Ahmad, S: A novel ubiquitous protein Chaperonin supports the endosymbiotic origin of mitochondrion and plant chloroplasts. *Biochem. Biophys. Res.* 163, 780-787 (1989).
- 39.- Hemmingsen, S.M., Ellis, R.J: Purification and properties of ribulose bisphosphate carboxylase large subunit binding protein. *Plant physiol.* 80: 269-276 (1986).
- 40.- Hemmingsen, S.M., Woolford, C., Van der vies, S.M., Tilly, K., Dennis, D.T., Geogopoulos, C.P, Hendrix, R.W., Ellis R.J: Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature.* 333, 330-334 (1988).
- 41.- Hendrick, J.P., Hartl, F.U: Molecular chaperon functions of heat shock proteins. *Ann.Rev.Biochem.* 62, 349-384 (1993).
- 42.- Horwich, A.L., Low, K.B., Fenton, W.A., Hirshfield, I.N., Furtak, K. *Cell*, 74, 909-917 (1993).

43. Higgins, D.G., Sharp, O: CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene*. 73, 237-244 (1988).
- 44.- Hubbard, T.J., Sander, C: The role of heat-shock and chaperone proteins in protein folding: Possible molecular mechanisms. *Protein Eng.* 4, 711-717 (1991).
- 45.- Jackson, G.S., Staniforth, R.A., Halsall, D.J., Atkinson, T., Holbrook, J.J., Clarke, A.R., Burston, S.G: Binding and hydrolysis of nucleotides in the chaperonin catalytic cycle: implications for the mechanism of assisted protein folding. *Biochemistry*. 32, 2554- 2563 (1993).
- 46.- Jenkins, A.J., March, J.B., Oliver, I.R., Masters, M: A DNA fragment containing the *groE* genes can suppress mutations in the *Escherichia coli dnaA* gene. *Mol Gen Genet*. 202. 446-454 (1986).
- 47.- Johnson, R.B., Fearon, K., Mason, T., Jindal, S: Cloning and characterization of the yeast chaperonin *HSP60* gene. *Gene*. 84, 295-302 (1989).
- 48.- Kang, P.J., Osterman, J., Shilling, J., Neupert, W., Craig, E.A., Pfanner, N: Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. *Nature*. 348: 137-143 (1990).
- 49.- Kyte, J., Doolittle, R.F: A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157, 105-132 (1982).
- 50.- Langer, T., Pfeifer, G., Martin, J., Baumeister, W., Hartl, F.U: Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO.J.*11, 4757-4765 (1992).

- 51.- Laskey, R.A., Honda, B.M., Mills, A.D., Finch, J.T: Nucleosomes are assembled by an acid protein which binds histones and transfers them to DNA. *Nature*. 275, 416-420 (1978).
- 52.- Lazcano, A.A., Fox, G.E., Oro, J.F: Life before DNA: The origin and evolution of early archean cells. In: R.P Mortlock ad., *The Evolution of Metabolic Function*. Boca Raton: CRC Press, pp 242-248 (1992).
- 53.- Lewis, V.A., Heynes, G.M., Zheng, D., Saibil, H., Wilson, K: T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature*. 358, 249-252 (1992).
- 54.- Lindquist, S., Craig, E.A: The Heat Shock Proteins. *Ann. Rev. Genet.* 22,631-677 (1988).
- 55.- Lissin, N.M., Venyaminov, S.Y., Girshovich, A.S: (Mg- ATP)-dependent self-assembly of molecular chaperone GroEL. *Nature*. 348: 339-342 (1990).
- 56.- Lubben, T.H., Gatenby, A.A., Donaldson, G.K., Lorimer, G.H, Viitanen, P.V: Identification of a groES-like chaperonin in mitochondria that facilitates protein folding. *Proc. Natl. Acad. Sci. USA*. 87: 7683-7687 (1990).
- 57.- Lubben, T.H., Donaldson, G.K., Viitanen, P.V., Gantenby, AA: Several proteins imported into chloroplasts from stable complexes with the GroEL-related chloroplast molecular chaperon. *Plant Cell*. 1: 1223-1230 (1989).
- 58.- Manning-Krieg, U.C., Scherer, P.E., Schatz, G: Sequential action of mitochondrial chaperone in protein import into the matrix. *EMBO. J*. 10: 3273-3280 (1991).

- 59.- Martel, R., Cloney, L.P., Pelcher, L.E., Hemmingsen, S.M: Unique composition of a plastid chaperonin-60: alfa y beta polipeptide- encoding genes are highly divergent. *Gene*. 94: 181-187 (1990).
- 60.- Martin, J., Mayhew, M., Langer, T., Hartl, F.U: The reaction cycle of GroEL and GroES in chaperonin-assisted protein folding. *Nature*. 366, 228-233 (1993).
- 61.- Martin, J., Langer, T., Boteva, R., Scramel, A., Horwich, A.L., Hartl, F.U: Chaperonin-mediated protein folding at the surface of groEL through a "molten globule" like intermediate. *Nature*. 352, 36-42 (1991).
- 62.- Martin, J., Geromanos, S., Tempst, P., Hartl, F.U: Identification of nucleotide binding regions in the chaperonin proteins GroEL and GroES. *Nature*. 366, 279-282 (1993).
- 63.- McMullin, T.W., Hallberg, R.L: A highly evolutionarily conserved mitochondrial protein is structurally related to the protein encoded by the *Escherichia coli* groEL gene. *Mol. Cell. Biol.* 8: 371-380 (1988).
- 64.- Mehra, V., Sweetser, D., Young, R.A: *Proc.Natl.Acad.Sci.Usa.* 83,7013-7017 (1986)
- 65.- Murakami, H., Pain, D., Blobel, G: 70-kD heat shock-related protein is one of at least two distinct cytosolic factors stimulating protein import into mitochondria. *J. Cell. Biol.* 107, 2051-2057 (1988).
- 66.- Musgrove, J.E., Jhonson, R.A., Ellis, R.J: Dissociation of the ribulose bisphosphate carboxilase large subunit binding protein into dissimilar subunits. *Eur. J. Biochem.* 163: 529-534 (1987).
- 67.- Myers, E.W., Miller, W. *CABIOS.* 4, 11-17 (1988).

- 68.- Normington, K., Khono, K., Kozutsumi, Y., Gething, M.J., Sambrook, J: *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. *Cell*. 57, 1223-1236 (1989).
- 69.- Novotny, J., Auffray, C. *Nucleic. Acid. Res.* 12, 243-255 (1984).
- 70.- Ostermann, J., Horwich, A.L., Neupert., Hartl, F. U: Protein. folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature*. 341: 125-130 (1989).
- 71.- Pelham, H.R.B: Heat shock and the sorting of luminal ER proteins. *EMBO.J.* 8. 3171-3176 (1989).
- 72.- Peralta, D., Hartman, D.J., McIntosh, A.M., Hoogenraad, N.J., Hoj, P.B: cDNA and deduced amino acid sequence of rat liver prehs60 (chaperonin-60). *Nucleic Acids Res.* 18, 7162-7167 (1990).
- 73.- Phipps, B.M., Typke, D., Hegerl, R., volker, s., Hoffmann, A., Stetter., Baumeister, W: Structure of a molecular from a Thermophilic archaebacterium. *Nature*. 361, 475-477 (1993).
- 74.- Pickets, D.J., Mayanil, Ch.S.K., Gupta, R.S: Molecular cloning of a chinese hamster mitochondrial protein related to the "chaperonin" family of bacterial and plant proteins. *J.Biol.Chem.* 264, 12001-12008 (1989).
- 75.- Prasad, T.K., Hack, E., Hallberg, R.L: Function of the maize mitochondrial chaperonin hsp60: specific association between hsp60 and newly synthesized F1-ATPase alpha subunits. *Mol. Cell. Biol.* 10: 3979-3986 (1990).

- 76.- Reading, D.S., Hallberg, R.L., Myers, A.M: Characterization of the yeast HSP60 gene coding for a mitochondrial assembly factor. *Nature*. 337: 655-659 (1989).
- 77.- Roy, H., Bloom, M., Milos, P., Monroe, M: Studies on the assembly of large subunits of ribulose biphosphate carboxylase in isolated pea chloroplasts. *J. Cell. Biol.* 94: 20-27 (1982).
- 78.- Rothman, J.E., Schmid, S.L: Enzymatic recycling of clathrin from coated vesicles. *Cell*. 46, 5-9 (1986).
- 79.- Saibil, H., Wood, S: Chaperonins. *Curr Opin Struct Biol.* 3, 207-213 (1993).
- 80.- Shinnick, T.M., Plikatis, B.P., Hycse, A.d., Van Landinham, R., Walker, L.L. *Nucleic Acids Res.* 17, 1254-1260 (1989).
- 81.- Taylor, W.R: The classification of amino acid conservation. *J.Theor.Biol.* 119, 205-218 (1986).
- 82.- Trent, J.D., Nimmegern, E., Wall, J.S., Hartl, F.U., Horwich, A.L: A molecular chaperone from a thermophilic archaeobacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature*. 354, 490-493 (1991).
- 83.- Van Dyk T.K., Gatenby, A.A., LaRosa, R.A: Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature*. 343: 451-453 (1989).
- 84.- Venner, T.J., Gupta, R.S: Nucleotide sequence of mouse HSP60 (chaperonin, GroEL homolog) cDNA. *Biochim.Biohys.Acta.* 1087, 336-338 (1990).

- 85.- Viitanen, P.V., Donaldson, G.K., Lorimer, H.L., Gatenby, A.A: Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. *Biochemistry*. 30, 9716-9723 (1991).
- 86.- Waldinger, D., Eckerskorn, C., Lottspeich, F., Cleve, H. *Biol.Chem.* 369, 1185-1189 (1988).
- 87.- Wallace, D.C: Structure and evolution of organelle genomes. *Microbiol. Rev.* 46, 208-240 (1982).
- 88.- Webb, R., Redy, K.J., Sherman, L.A: regulation and sequence of the *Synechococcus* sp. Strain PCC 7942 *groESL* operon, encoding a cyanobacterial chaperonin. *J. Bact.* 172, 5079-5088 (1990).
- 89.- Woese, C.R: Bacterial evolution. *Microbiol Rev.* 51, 221-271 (1987).
- 90.- Woese, C.R., Gibson, J., Fox, G.E: Do genealogical patterns in purple photosynthetic bacteria reflect interspecific gene transfer?. *Nature*. 283, 212-214 (1980).
- 91.- Woese, C.R., Kandler, O., Wheelis, M.L: Towards a natural system of organisms: Proposal for the domains archea, bacteria, and eucarya. *Proc.Natl.Acad.Sci. USA.* 87, 4576-4579 (1990).
- 92.- Yaffe, M.B., Farr, G.W., Milkos, D., Horwich, A.L., Sternlicht, M.L., Sternlicht, H: TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature*. 358, 245-248 (1992).
- 93.- Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., Woese, C.R: Mitochondrial origins. *Proc. Natl. Acad. Sci. USA.* 82. 4443-4447 (1985).

94.- Zabaleta, E., Oropeza, A., Jiménez, B., Salerno, G., Crespi, M., Herrera-Estrella, L: Isolation and characterization of genes encoding chaperonin 60Beta from *Arabidopsis thaliana*. *Gene*. 111, 175-181 (1992).

95.- Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C: The universally conserved GroE (Hsp60) chaperonins. *Annu. Rev. Microbiol.* 45: 301-325 (1991).

96.- Zvelebil, M.J., Barton, G.J., Taylor, W.R., Sternerberg, W: Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* 195, 957-961 (1987).

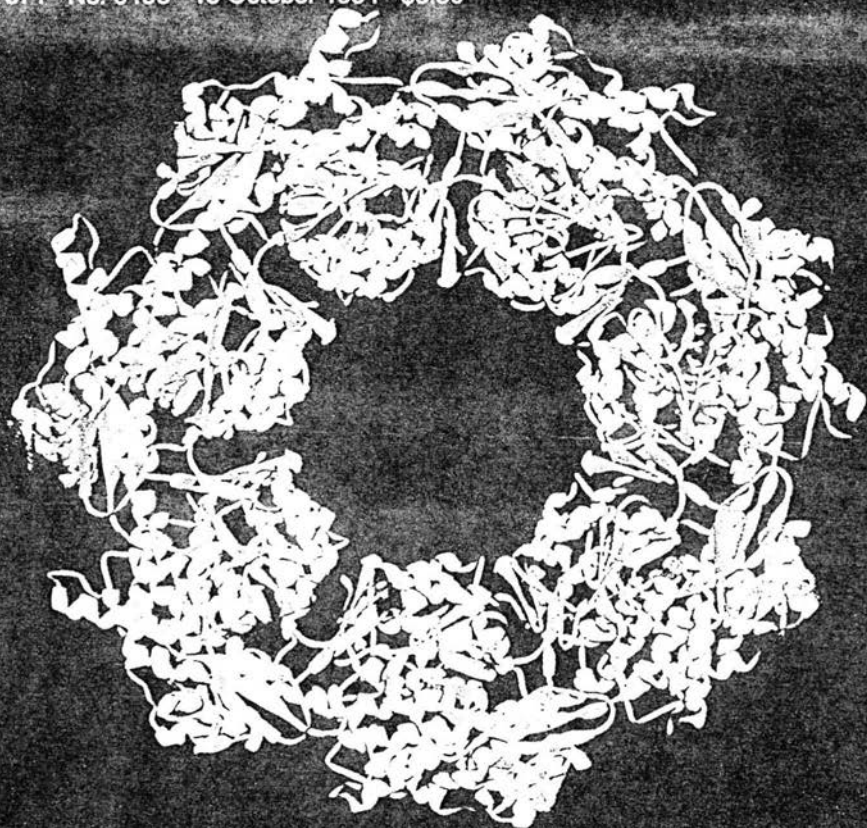
A P E N D I C E

Una vez concluido este trabajo, en la revista NATURE, No 6498, Vol 371 del 13 de Octubre de 1994, aparecieron una serie de artículos dedicados a la estructura y función de GroEL, en ellos se describe con cierto detalle la estructura tridimensional y los dominios funcionales del péptido. Aquí reproducimos algunos aspectos importantes de esos trabajos.

nature

INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Volume 371 No. 6498 13 October 1994 \$8.50



GroEL structure and function

The crystal structure of the bacterial chaperonin GroEL at 2.8 Å

Kerstin Braig^{*†}, Zbyszek Otwinowski^{†‡§}, Rashmi Hegde^{†‡§}, David C. Boisvert^{*}, Andrzej Joachimiak^{‡§}, Arthur L. Horwich^{*†} & Paul B. Sigler^{†‡||}

* Department of Genetics and † Department of Molecular Biophysics and Biochemistry and ‡ Howard Hughes Medical Institute, Yale University School of Medicine, Boyer Center, 295 Congress Avenue, New Haven, Connecticut 06510, USA

The crystal structure of *Escherichia coli* GroEL shows a porous cylinder of 14 subunits made of two nearly 7-fold rotationally symmetrical rings stacked back-to-back with dyad symmetry. The subunits consist of three domains: a large equatorial domain that forms the foundation of the assembly at its waist and holds the rings together; a large loosely structured apical domain that forms the ends of the cylinder; and a small slender intermediate domain that connects the two, creating side windows. The three-dimensional structure places most of the mutationally defined functional sites on the channel walls and its outward invaginations, and at the ends of the cylinder.

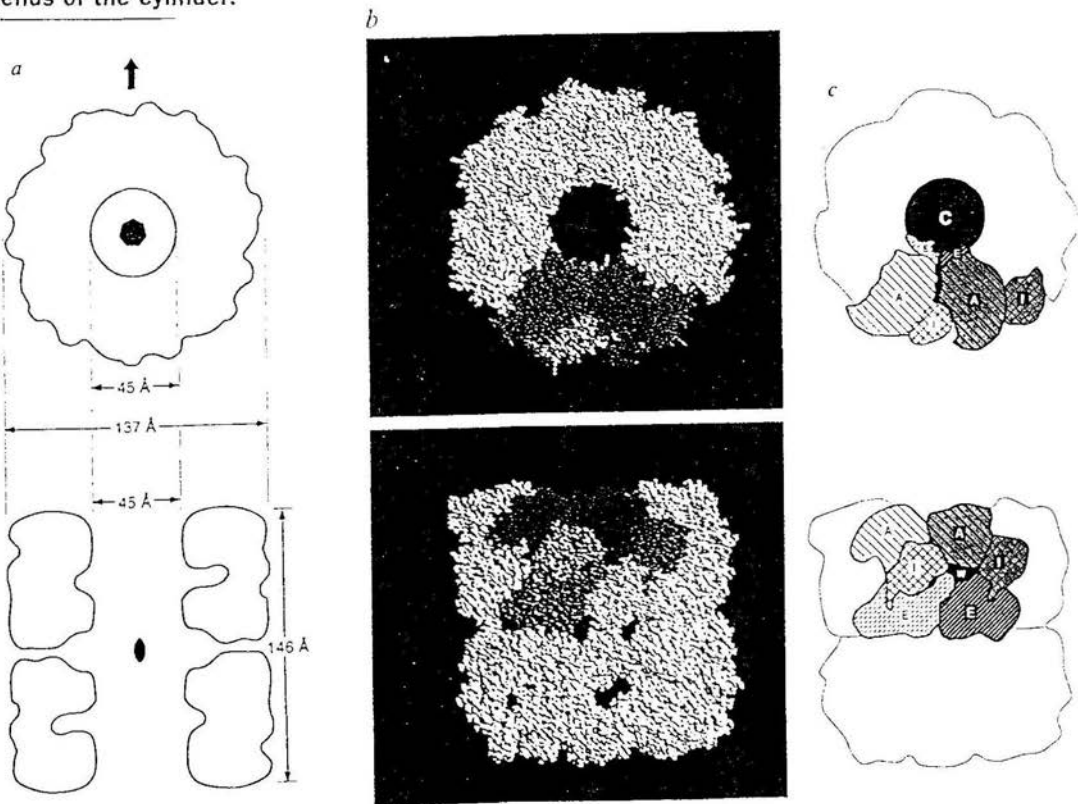


FIG. 1. Architecture of GroEL. *a*, Principal dimensions on a schematic diagram of the double-ring cylinder, showing a view into the central channel (top panel), and a vertical cut through the cylinder made by a plane defined by a diameter and the cylindrical axis (bottom panel). The dimensions are rough because the surfaces are irregular. The walls of the inner channel contain substantial bulges at the level of the intermediate domain which are not shown. *b*, Van der Waals space-filling model of the entire 14-mer shown in grey except for two adjacent subunits in the 'top' ring seen facing into the channel (top panel) and

viewed from the outside (bottom panel). Domains within the left subunit are colour-coded as follows: equatorial, green; intermediate, gold; and apical, purple; the right subunit's equatorial, intermediate and apical domains are colour-coded yellow, red and blue, respectively (see Figs 2, 3 and text for discussion of domains). *c*, Drawing depicting diagrammatically the structure in *b*. E, I and A, equatorial, intermediate and apical domains, respectively; C, central channel; W, external opening of a side window.

Residues in chaperonin GroEL required for polypeptide binding and release

Wayne A. Fenton, Yechezkel Kashi*,
Krystyna Furtak† & Arthur L. Horwich†‡

Department of Genetics and † Howard Hughes Medical Institute,
Yale University School of Medicine, Boyer Center,
295 Congress Avenue, New Haven, Connecticut 06510, USA

* Present address: Department of Biotechnology, Technion Institute,
Haifa 32000, Israel

‡ To whom correspondence should be addressed

CHAPERONINS are ring-shaped protein complexes that are essential in the cell, mediating ATP-dependent polypeptide folding in a variety of compartments¹⁻³. Recent studies suggest that they function through multiple rounds of binding and release of non-native proteins; with each round of ATP-driven release into the bulk solution, a substrate protein kinetically partitions between folding to the

native state or rebinding to another chaperonin molecule^{4,5}. To gain further insight into the mechanism of polypeptide binding and release by the chaperonin GroEL from *Escherichia coli*, we have undertaken a mutational analysis that relates the functional properties of GroEL to its crystal structure⁷. Our functional tests identify a putative polypeptide-binding site on the inside surface of the apical domain, facing the central channel, consisting of hydrophobic residues. These same residues are essential for binding of the co-chaperonin GroES, which is required for productive polypeptide release. A highly conserved residue, Asp87, positioned within a putative nucleotide-binding pocket in the top of the equatorial domain, is essential for ATP hydrolysis and polypeptide release.

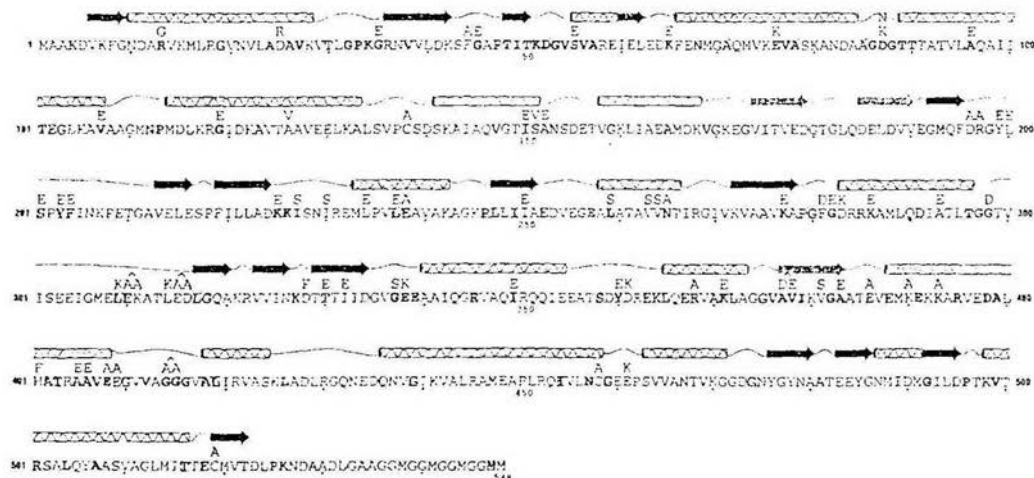


FIG. 1 Amino-acid substitutions in GroEL are shown above the wild-type amino-acid sequence. Secondary structure⁷ (coil for α -helix, arrow for β -sheet) is indicated by the overline, coloured according to domain: equatorial, green; intermediate, blue; apical, red. Residues conserved in both Hsp60/GroEL and TF55/TCP1 chaperonin families are shown in bold. Residues identical in at least 49 out of 50 compared Hsp60/GroEL sequences are shown on orange background.

METHODS. Fragments bearing site-directed mutations were produced

by PCR using as template a wild-type GroE operon, and substituted for the wild-type sequence in a *trc*-expression plasmid. For each mutation, a restriction fragment bearing only the appropriate mutation as confirmed by sequence analysis, was recloned. All mutations are single-residue substitutions except for the following double mutants: R13G/A126V, G45E/I292T, E310A/K311A, E315A/D316A, G337S/I349E, A377D/V378E, E386A/K390A, K390A/K393A and G414A/G415A.